

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 7, 2002, 18:47:47 ; Search time 1988.65 Seconds  
(without alignments)  
8552.833 Million cell updates/sec  
Title: US-09-665-728-2  
Perfect score: 1031  
Sequence: 1 tctagcaacccttcggcc.....agccggctcgagggccgc 1031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.om.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: gb.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1031	100.0	1031	6	AX108634	AX108634 Sequence
2	336.8	32.7	18207	9	AF156100	AF156100 Homo sapi
3	305.8	29.7	191750	2	AC024113	AC024113 Mus muscu
c 4	215.8	20.9	191750	2	AC024113	AC024113 Mus muscu
5	215.8	20.9	230755	2	AC024618	AC024618 Mus muscu
6	185.2	18.0	15597	3	AF074901	AF074901 Caenorhab
7	183.4	17.8	182211	9	AL360004	AF074901 Caenorhab
8	108.4	10.5	41345	3	CEB15G9	AF074901 Caenorhab
9	91.6	9.1	1610	9	HS245418	AF074901 Caenorhab
10	91.4	8.9	4312	10	AF134319	AF134319 Homo sapi
11	89.4	8.7	157029	9	AL135796	AF134319 Homo sapi
12	69.8	6.8	87461	9	HS218G19	AF135796 Human DNA
c 13	69.8	6.8	180248	9	AC023275	AF135796 Human DNA
c 14	65.4	6.3	717	11	HSJ34G10	AC023275 Homo sapi
15	61.6	6.0	356	11	HSJ34G10	AL159308 STS from
16	51.6	5.0	1873	3	NCU37520	Z51502 H.sapiens (
17	50.4	4.9	2336	3	NEPDSF	U37520 Nephila cla
18	50.4	4.9	2338	6	AR088543	M37137 N.clavipes
19	50.4	4.9	2338	6	I92789	AR088543 Sequence
20	50.2	4.9	4359	10	D85391	I92789 Sequence 1
21	48.6	4.7	87810	2	AC022987	D85391 Mus musculu
22	47.6	4.6	154137	8	AP003215	AC022987 Homo sapi
c 23	47	4.6	41173	1	SCI41	AP003215 Oryza sat
c 24	46.2	4.5	3867	1	AY033407	AL132648 Streptomy
25	46	4.5	1026	3	AF350272	AF003215 Oryza sat
26	44.8	4.3	43349	1	SCR41	AL132648 Streptomy
27	44.6	4.3	4377	10	RND62897	AF350272 Gasteraca
c 28	44.2	4.3	44109	1	SCSF1	AL117387 Streptomy
29	44	4.3	12606	1	AE005995	U62897 Rattus norv
c 30	43.8	4.2	18023	1	AX024392	AL450165 Streptomy
c 31	43.8	4.2	18023	6	AX024285	AE005995 Caulobact
c 32	43.8	4.2	87810	2	AC022987	AX024392 Sequence
33	43.4	4.2	2696	1	AF083621	AX024285 Sequence
34	43.4	4.2	2745	6	AX179756	AC022987 Homo sapi
35	43	4.2	2047	10	BC003909	AF083621 Lysobacte
36	43	4.2	5991	10	AF199422	AX179756 Sequence
37	43	4.2	6423	10	AF199421	BC003909 Mus muscu
c 38	43	4.2	198677	1	AE001863	AF199421 Mus muscu
39	42.6	4.1	132470	9	HS203P18	AE001863 Deinococc
40	42.4	4.1	3203	10	AF008561	297180 Human DNA s
41	42.4	4.1	12356	10	AF134318	AF008561 Mus muscu
42	42.4	4.1	135545	10	MMHC213L3	AF134318 Mus muscu
c 43	42.4	4.1	224020	2	AC087117	AF109905 Mus muscu
44	42.2	4.1	100000	9	AP000503	AC087117 Mus muscu
45	42.2	4.1	125350	2	AC020768	AP000503 Homo sapi
						AC020768 Homo sapi

ALIGNMENTS

RESULT 1	AX108634	Sequence 2 from Patent WO0123419.	1031 bp	DNA	PAT	30-APR-2001
AX108634	Sequence 2 from Patent WO0123419.					
LOCUS	AX108634					
DEFINITION	Sequence 2 from Patent WO0123419.					
ACCESSION	AX108634					
VERSION	AX108634.1	GI:13923866				
KEYWORDS	Norway rat.					
SOURCE	Rattus norvegicus					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
REFERENCE	1 (bases 1 to 1031)					
AUTHORS	Stanton, L.W. and Kapoun, A.M.					
TITLE	Differentially expressed genes					
JOURNAL	Patent: WO 0123419-A 2 05-APR-2001;					
FEATURES	SCIOS INC. (US)					
source	Location/Qualifiers					
	1..1031					
	/organism="Rattus norvegicus"					

[illegible]



## AUTHORS

Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,  
Dederich, D., Thomas, S., Okunolu, G., Carlock, C., Garner, T.,  
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
Fernandez, C., Ferraruto, D., Forcum-Tansey, J., Gill, R.,  
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,  
Hosack, H., Hou, X., Huber, J., Jackson, L., Jia, X., Kelly, J., Kelly, S.,  
Kovar, C., Liu, J., Liu, W., Loulsegé, H., Lozardo, R.J., Martin, R.,  
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,  
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Wallington, S.,  
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,  
Worley, K. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 191750)  
Worley K.C.  
Direct Submission  
Submitted (24-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jan 4, 2001 this sequence version replaced gi:11096415.

COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: MABZ  
Center clone name: RP23-277F6  
----- Summary Statistics  
Sequencing vector: ML3; L08821  
Chemistry: Dye-primed Bodypy; 74% of reads  
Chemistry: Dye-terminator Big Dye; 26% of reads  
Assembly: Program: Phrap; version 0.990329  
Consensus quality: 159536 bases at least Q40  
Consensus quality: 173438 bases at least Q30  
Consensus quality: 179682 bases at least Q20  
Estimated insert size: 180449; sum-of-contrigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 3.2x in Q20 bases; sum-of-contrigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 29 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

[illegible]

* 130074	130173:	gap of unknown length
* 130174	135444:	contig of 5171 bp in length
* 135445	135444:	gap of unknown length
* 135446	140753:	contig of 5309 bp in length
* 140754	140853:	gap of unknown length
* 140854	146305:	contig of 5452 bp in length
* 146306	146405:	gap of unknown length
* 148406	150873:	contig of 4468 bp in length
* 150874	150973:	gap of unknown length
* 150974	155410:	contig of 4437 bp in length
* 155411	155411:	gap of unknown length
* 155511	155684:	contig of 4174 bp in length
* 159685	159784:	gap of unknown length
* 159785	163180:	contig of 3396 bp in length
* 163181	163280:	gap of unknown length
* 163281	163459:	contig of 2179 bp in length
* 165460	165559:	gap of unknown length
* 165560	168279:	contig of 2720 bp in length
* 168280	168379:	gap of unknown length
* 168380	170847:	contig of 2468 bp in length
* 170848	170947:	gap of unknown length
* 170948	175115:	contig of 4168 bp in length
* 175116	175216:	gap of unknown length
* 175216	177870:	contig of 2655 bp in length
* 177871	177970:	gap of unknown length
* 177971	181412:	contig of 3442 bp in length
* 181413	181512:	gap of unknown length
* 181513	183055:	contig of 1543 bp in length
* 183056	183155:	gap of unknown length
* 183156	184937:	contig of 1782 bp in length
* 184938	185037:	gap of unknown length
* 185038	187608:	contig of 2571 bp in length
* 187609	187708:	gap of unknown length
* 187709	189220:	contig of 1512 bp in length
* 189221	189320:	gap of unknown length
* 189321	190396:	contig of 1076 bp in length
* 190397	190496:	gap of unknown length
* 190497	191750:	contig of 1254 bp in length.
FEATURES		
Location/Qualifiers		
1..191750		
/organism="Mus musculus"		
/db_xref="taxon:10090"		
/chromosome="11"		
/clone="RP23-277F6"		
BASE COUNT	47260 a	47982 c 47655 g 46018 t 2835 others
ORIGIN	.	
Query Match	29.7%;	Score 305.8; DB 2; Length 191750;
Best Local Similarity	92.3%;	Pred. No. 4.8e-52;
Matches 322; Conservative	0; Mismatches 27; Indels 0; Gaps	
Qy 19	ccgcctagacgagactgaactgccatcatcctcgacctgcgcgtccccattagggcct 78	
Db 191108	CCGCCTAGACGGACTGCACTGCCATCTCCTACTCGACCCTCGGCTCCCATCAGGCT 191167	
Qy 79	gcagcttcgggctcaagataagccttagggcgagctcctcgctgcctcctggcgacct 138	
Db 191168	GCAGCCTTCCGGCTCACACTGACGCCCTTGSGCGCAGCTCTGTCCGATGTCGTGGCGATCT 191227	
Qy 139	acacagtatgcccgcgcggcgtcacatctgatgaccacgaagcgcgtgccccgcaca 198	
Db 191228	CTACAGCAGTGGCGGGGTGGTCCAGCTCTGACGCCGCCACGAGAGCGCTGTCCC CGCCA 191287	
Qy 199	caggagacgcacccctggccttcttgatttcgatgttacccggctccatttgtggagcatctga 258	
Db 191288	CAGGAGACGCCACCTAGCCTTCTGATCTCATCCGGCTCCATFTGGGATGACCTGA 191347	
Qy 259	tgaagtgtatcgacggcgccctaagcattcttgaggcagcagctgtgaacacagcacgccggg 318	
Db 191348	TGCAGTGTATCGACGGCGCCCTCGCGCATTTCTGGAGCGCAGCTTGACACCCGACGCCGG 191407	
Qy 319	tcatcgccaactatggcctggtgcctttccacgaccacagacattggccc 367	



```

|||||
Db 191408 TCATCCCAACATGCGCTGGTGGCTTTCCACAGCCAGGTAGAGCCCC 191456
|||||

RESULT 4
AC024113/c
LOCUS
DEFINITION
AC024113 191750 bp DNA HTG 07-JAN-2001
SEQUENCE, 29 unordered pieces.
AC024113.12 GI:12025591
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 191750)
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, X., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hoques, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogum, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Wozny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Murray, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 191750)
Worley, K.C.
Direct Submission
Submitted (24-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:11096415.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAB2
Center clone name: RP23-277F6
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 74% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159536 bases at least Q40
Consensus quality: 173428 bases at least Q30
Estimated insert size: 180449; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 19880: contig of 19880 bp in length
* 19881 19980: gap of unknown length

```

```

19981 38620: contig of 18640 bp in length
38621 38720: gap of unknown length
38721 51091: contig of 12371 bp in length
51092 51191: gap of unknown length
51192 63586: contig of 12395 bp in length
63587 63686: gap of unknown length
63687 76765: contig of 13079 bp in length
76766 76866: gap of unknown length
76867 92384: contig of 15519 bp in length
92385 92484: gap of unknown length
92485 105754: contig of 13270 bp in length
105755 105854: gap of unknown length
105855 114346: contig of 8492 bp in length
114347 114446: gap of unknown length
114447 122760: contig of 8314 bp in length
122761 122860: gap of unknown length
122861 130073: contig of 7213 bp in length
130074 130173: gap of unknown length
130174 133344: contig of 5171 bp in length
133345 135444: gap of unknown length
135445 140753: contig of 5309 bp in length
140754 140853: gap of unknown length
140854 146305: contig of 5452 bp in length
146306 146405: gap of unknown length
146406 150873: contig of 4468 bp in length
150874 150973: gap of unknown length
150974 155410: contig of 4437 bp in length
155411 155510: gap of unknown length
155512 159684: contig of 4174 bp in length
159685 159784: gap of unknown length
159785 163180: contig of 3396 bp in length
163181 163280: gap of unknown length
163281 165450: contig of 2179 bp in length
165451 165559: gap of unknown length
165560 168279: contig of 2720 bp in length
168280 168379: gap of unknown length
168380 170847: contig of 2468 bp in length
170848 170947: gap of unknown length
170949 175115: contig of 4168 bp in length
175116 175215: gap of unknown length
175216 177870: contig of 2655 bp in length
177871 177970: gap of unknown length
177971 181412: contig of 3442 bp in length
181413 181512: gap of unknown length
181513 183055: contig of 1543 bp in length
183056 183155: gap of unknown length
183156 184937: contig of 1782 bp in length
184938 185037: gap of unknown length
185038 187608: contig of 2571 bp in length
187609 187708: gap of unknown length
187709 189220: contig of 1512 bp in length
189221 189320: gap of unknown length
189321 190396: contig of 1076 bp in length
190397 190496: gap of unknown length
190497 191750: contig of 1254 bp in length.
FEATURES
Location/Qualifiers
source
1. 191750
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-277F6"
/chromosome="11"

```

```

BASE COUNT 47360 a 47982 c 47655 g 46018 t 2835 others
ORIGIN

```

```

Query Match 20.9%; Score 215.8; DB 2; Length 191750;
Best Local Similarity 84.8%; Pred. No. 4.9e-34;
Matches 267; Conservative 0; Mismatches 42; Indels 6; Gaps 2;

```

```

QY 709 aggtgttaagtgaggtagtcgcgcacatccagcgcctccaaagttcactcgtctcagcag 768
|||||
Db 13539 AGGTGTTAAAGTGGGTGGAGTGCAGCAATCCAGGCTCCAAAGTTTCATCTGCTGTCAG 13480

```







```

QY 614 cggtagcgcaccaccctggctacgtggttttggagagatgcctccaccagttctgg 673
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 TGGTAATAGACACATCCAGGATTCGTACATATGAAAAAATCGCCGCTGCACTTTTGG 569
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 674 ccaagtgttcagctgacagcagcagcagcagcagcagcagcagcagcagcagc 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 570 ACAAGTTTTCATCTTGAAGAAGAGTGTATGATGACGTATGATGATGATGATGATG 629
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 734 catccagcctccaaattcatctgtcagcagacacacagcagcagcagcagcagc 793
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 TGTCAACAGAGAAGAAAGTTTCATTTGATGACGAGGCACGTGAACGGGAGGAGACAGTGT 689
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 794 atggagaatcctttgacccacgcttgagggaagtcaccatctcactgaagcgg 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 TAGAATATCCCATGACACAGCATCTCTCAGAGCTCACCATCTCATTTGTGAGG 743

RESULT 7
LOCUS AL360004 182211 bp DNA PRI 16-AUG-2001
DEFINITION Human DNA sequence from clone RP11-88G17 on chromosome 9, complete
sequence.
ACCESSION AL360004
VERSION AL360004.22 GI:15212049
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182211)
AUTHORS Kimberley, A.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Aug 18, 2001 this sequence version replaced gi:15020463.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-88G17 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-88G17 The true
right end of clone RP11-202H3 is at 75398 in this sequence.
Location/Qualifiers
1..182211
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-88G17"
/clone_lib="RPC1-11.1"
misc_feature 43597..44209

```

```

/note="Sequence from AC006241 sequenced by WIBR"
50279..50351
/note="Sequence from AC006241 sequenced by WIBR"
93314
/note="Tandem repeat. Forced join. Gap sized to be
approximately 150bp from restriction digest data and
spanning pUC clone."
BASE COUNT 40368 a 47059 c 48127 g 46657 t
ORIGIN

Query Match 17.8%; Score 183.4; DB 9; Length 182211;
Best Local Similarity 77.4%; Pred. NO. 1.5e-27;
Matches 236; Conservative 0; Mismatches 66; Indels 3; Gaps 1;

QY 80 cagcctcggctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137293 CAGAGCCGCGTCGCGCATGATGATGATGATGATGATGATGATGATGATGATG 137352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 140 cacagtagtgggcggcggcggcggcggcggcggcggcggcggcggcggcggc 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137353 TCGCGCAGTGGCAGTGGCAGTGGC---GGGGCGCGCGGACGTAATGCCCCAC 137409
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 200 aggagagcaccctggccttcgttcgttcgttcgttcgttcgttcgttcgttcgt 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137410 GGGGAGCGCCACCCTGGCTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 137469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 gcaagtatgacagcggcggcggcggcggcggcggcggcggcggcggcggcgg 319
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137470 GCAGGTGATGCATGGCGCTCGCGCATCTCTGGAACGACGTAATGCCCGCAGCC 137529
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 320 catcgcaactatgcgtggtgcttcacacgacacacacacacacacacacacacac 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137530 CATGCCCAACTACGGCTGGTGGCTTCACAGCACCAGGAGGAGGAGGAGGAGG 137589
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 ggcgg 384
      |||
Db 137590 GCGCG 137594

RESULT 8
LOCUS CEF15G9 41345 bp DNA INV 20-JUN-2001
DEFINITION Caenorhabditis elegans cosmid F15G9, complete sequence.
ACCESSION Z47068
VERSION Z47068.1 GI:634018
KEYWORDS HTG; Fibronectin type III repeat; Transfer-RNA; tRNA-Ala.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 41345)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 41345)
AUTHORS Sulston, J.E.
TITLE Direct Submission
JOURNAL Submitted (24-DEC-1994) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
COMMENT

```

IMPORTANT: This sequence is not the entire insert of clone F15G9. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.  
The true left end of clone F15G9 is at 1 in this sequence. The true right end of clone F15G9 is at 4082 in sequence 247070.  
The true left end of clone T0989 is at 41245 in this sequence. The true right end of clone C43C3 is at 8835 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 247067.  
The end of this sequence: (41245..41345) overlaps with the start of sequence 247070.  
For a graphical representation of this sequence and its analysis see:- <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F15G9>.

FEATURES  
source

Location/Qualifiers  
1..41345  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="X"  
/clone="F15G9"  
Join(2873..3089,3208..3273,3480..3823,3869..3947,4131..4264)  
/gene="F15G9.1a"  
Join(2873..3089,3208..3273,3480..3823,3869..3947,4131..4264)  
/gene="F15G9.1a"  
/note="cDNA EST yk102e12.3 comes from this gene  
cDNA EST yk102e12.5 comes from this gene  
cDNA EST yk168h5.3 comes from this gene  
cDNA EST yk168h5.5 comes from this gene  
cDNA EST yk450c6.3 comes from this gene  
cDNA EST yk450c6.5 comes from this gene  
cDNA EST yk268g4.3 comes from this gene  
cDNA EST yk268g4.5 comes from this gene  
cDNA EST yk252c2.3 comes from this gene  
cDNA EST yk252c2.5 comes from this gene  
cDNA EST yk527d8.3 comes from this gene  
cDNA EST yk558g5.3 comes from this gene  
cDNA EST yk584g11.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAA87331.1"  
/db\_xref="GI:3875951"  
/db\_xref="SWISS-PROT:Q10033"  
/translation="MSSRTSYTPSPSRKYGWDYSSYSSTSRDRDTSYRDRDY  
SSTYSITRPRYSTAYSKYSTDLGKTDGMDYSAAPIAKDEIEELSDAKTPTTPT  
EETVDAPREVLQESQISDFVEEDDDQTEVEEVLKLTLASPEESEEEDDEDVRKAE  
LLAASTQIRESSPVSPKVEYSSAALFANDNGNETNRTSPPTVLNAKLGIGIPWP  
KSSDVKKEGDAEALPKRYSDLLIARENTGVVEESKTTDDAYKNEYGAGTNGVKVSTH  
NFA"  
Join(3213..3273,3480..3823,3869..3947,4131..4264)  
/gene="F15G9.1b"  
Join(3213..3273,3480..3823,3869..3947,4131..4264)  
/gene="F15G9.1b"  
/note="cDNA EST yk527d8.5 comes from this gene  
cDNA EST yk558g5.5 comes from this gene  
cDNA EST yk584g11.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB54215.1"  
/db\_xref="GI:5824447"  
/db\_xref="SPTREMBL:O9U311"  
/translation="MDYSAAPTAKDAIEELSDAKTPTTPTETVDAPREVLQESQI  
SDFVEEDDDQTEVEEVLKLTLASPEESEEEDDEDVRKAEALLAASTQIRESSPVSS  
PVEVSSAALFANDNGNETNRTSPPTVLNAKLGIGIPWPKSSDVKKEGDAEALP  
KKRVSOLIARENTGVVEESKTTDDAYKNEYGAGTNGVKVSTHFA"  
complement(Join(6076..6174,6498..6562,6959..7029,  
7077..7256,7305..7378))  
/gene="F15G9.3"  
complement(Join(6076..6174,6498..6562,6959..7029,  
7077..7256,7305..7378))  
/gene="F15G9.3"  
/codon\_start=1

/protein\_id="CAA87333.1"  
/db\_xref="GI:3875953"  
/db\_xref="SWISS-PROT:Q10035"  
/translation="MTHQPTLYQLKKQKQYIFLKFNSQLFWYIMFSTNDNAKYS  
KIDFFHVLLOFYAVKQTIHQHQRSGSDPDWDEFFRIYSADKRPIDRINARQCHSAGI  
KLNLNKLDQKLNFTDYDSINDKLTPTYAMSLWLNNSFFPSYLNQAGKAYAPMTKSD  
SE"  
complement(Join(8996..9122,9175..9413,9464..9559))  
/gene="F15G9.2"  
complement(Join(8996..9122,9175..9413,9464..9559))  
/gene="F15G9.2"  
/codon\_start=1  
/protein\_id="CAA87332.1"  
/db\_xref="GI:3875952"  
/db\_xref="SWISS-PROT:Q10034"  
/translation="MDFNSLFLNSKIEDSDNDYRNSDCLSERPKKTSKTRSIKWS  
ESDLFKMSKIDYAVETMRNRNLEEVFSNYSYFILLROVSOQMDRLNAGTENILETG  
EVSDELSKKKTFVLMNPTAVLKVMTPDPSNDGVYRSLRFGMWYLR"  
Join(13472..13724,14222..14326,15572..16215,16297..16402,  
16485..16806,16959..17346,17399..17570,17654..18468,  
18624..18747,18910..19133,19183..19581,19759..19881,  
19935..20717,20970..21059,21111..21198,21257..21441,  
21488..21590,23087..23202,23255..23363,23492..23693,  
23780..23939,23988..24077,24544..25504,25673..25874,  
25934..26313,26565..26955,27233..27476,27750..28067,  
28340..28517,28564..28842,28890..29002,36561..36632,  
36681..36762,37047..37126,38336..38407,38452..38575,  
38622..38773,38829..38952,39009..39082,39698..39828,  
39875..39939,40081..40209,40328..41345,247070.1:102..304,  
247070.1:350..522,247070.1:679..802,247070.1:1753..1890,  
247070.1:1941..2081,247070.1:2213..2770,  
247070.1:2834..3265,247070.1:3313..3426,  
247070.1:3882..4289,247070.1:4333..4473,  
247070.1:4523..4666,247070.1:4712..4837,  
247070.1:4944..5392,247070.1:5571..6141,  
247070.1:6348..6539,247070.1:6589..6774,  
247070.1:7235..7444,247070.1:7490..7687,  
247070.1:7828..7926,247070.1:7972..8171)  
/gene="F15G9.4b"  
Join(13472..13724,14222..14326,15572..16215,16297..16402,  
16485..16806,16959..17346,17399..17570,17654..18468,  
18624..18747,18910..19133,19183..19581,19759..19881,  
19935..20717,20970..21059,21111..21198,21257..21441,  
21488..21590,23087..23202,23255..23363,23492..23693,  
23780..23939,23988..24077,24544..25504,25673..25874,  
25934..26313,26565..26955,27233..27476,27750..28067,  
28340..28517,28564..28842,28890..29002,36561..36632,  
36681..36762,37047..37126,38336..38407,38452..38575,  
38622..38773,38829..38952,39009..39082,39698..39828,  
39875..39939,40081..40209,40328..41345,247070.1:102..304,  
247070.1:350..522,247070.1:679..802,247070.1:1753..1890,  
247070.1:1941..2081,247070.1:2213..2770,  
247070.1:2834..3265,247070.1:3313..3426,  
247070.1:3882..4289,247070.1:4333..4473,  
247070.1:4523..4666,247070.1:4712..4837,  
247070.1:4944..5392,247070.1:5571..6141,  
247070.1:6417..6539,247070.1:6589..6774,  
247070.1:7235..7444,247070.1:7490..7687,  
247070.1:7828..7926,247070.1:7972..8171)  
/gene="F15G9.4a"  
Join(13472..13724,14222..14326,15572..16215,16297..16402,  
16485..16806,16959..17346,17399..17570,17654..18468,  
18624..18747,18910..19133,19183..19581,19759..19881,  
19935..20717,20970..21059,21111..21198,21257..21441,  
21488..21590,23087..23202,23255..23363,23492..23693,  
23780..23939,23988..24077,24544..25504,25673..25874,  
25934..26313,26565..26955,27233..27476,27750..28067,  
28340..28517,28564..28842,28890..29002,36561..36632,  
36681..36762,37047..37126,38336..38407,38452..38575,  
38622..38773,38829..38952,39009..39082,39698..39828,  
39875..39939,40081..40209,40328..41345,247070.1:102..304,  
247070.1:350..522,247070.1:679..802,247070.1:1753..1890,  
247070.1:1941..2081,247070.1:2213..2770,  
247070.1:2834..3265,247070.1:3313..3426,  
247070.1:3882..4289,247070.1:4333..4473,  
247070.1:4523..4666,247070.1:4712..4837,  
247070.1:4944..5392,247070.1:5571..6141,  
247070.1:6417..6539,247070.1:6589..6774,  
247070.1:7235..7444,247070.1:7490..7687,  
247070.1:7828..7926,247070.1:7972..8171)  
/gene="F15G9.4a"  
Join(13472..13724,14222..14326,15572..16215,16297..16402,  
16485..16806,16959..17346,17399..17570,17654..18468,  
18624..18747,18910..19133,19183..19581,19759..19881,  
19935..20717,20970..21059,21111..21198,21257..21441,  
21488..21590,23087..23202,23255..23363,23492..23693,  
23780..23939,23988..24077,24544..25504,25673..25874,  
25934..26313,26565..26955,27233..27476,27750..28067,  
28340..28517,28564..28842,28890..29002,36561..36632,  
36681..36762,37047..37126,38336..38407,38452..38575,  
38622..38773,38829..38952,39009..39082,39698..39828,  
39875..39939,40081..40209,40328..41345,247070.1:102..304,  
247070.1:350..522,247070.1:679..802,247070.1:1753..1890,  
247070.1:1941..2081,247070.1:2213..2770,

247070.1:2834. 3265,247070.1:3313. 3426,  
247070.1:3882. 4289,247070.1:4333. 4473,  
247070.1:4523. 4666,247070.1:4712. 4837,  
247070.1:4944. 5392,247070.1:5571. 6141,  
247070.1:6348. 6539,247070.1:6589. 6774,  
247070.1:7235. 7444,247070.1:7490. 7687,  
247070.1:7828. 7926,247070.1:7972. 8171)  
/gene="F15G9.4a"  
/note="contains similarity to Pfam domain: PF00047  
(Immunoglobulin domain), Score=1411.6, E-value=0, N=47  
cDNA EST YK556.f1.5 comes from this gene"  
/codon\_start=-1  
/protein\_id="CAA87336.1"  
/db\_xref="GI:3875956"  
/db\_xref="SPTREMBL:O76518"  
/translation="MGSPSWLYGVGLLLATTCSSYNDKNDKPTGKSSLAFFDIT  
GSFMDLVQVREGAAIKFTVMAQREKLIYNIIVFPDHPYLGEIINTDSTFMKOL  
SVYVHTGDCGNTHPFRFYKIAAASFGVFKLEKSDVSTVLEKRVHAKVKKVHL  
MYEARERGTVSRNIPVDKHLSELTISLQDKDSDNLDIVLRDPGRVTKLYSKE  
GSTIDLNKVLIRLKPSPGVWTVNTNSRLKHTIRVFGHCAVDKYGFAFASRLDRLEL  
ARRPVLNQDTYLLINMTGLIPGTGVEIDLYIHGSLYKAVASPHRTPNMYFAGP  
FVPPKGLFVVRVQGYDNDYEFMRIAFTAGTSVIVGPFRAFMSPIHQEFVGRDLNLS

Query Match 10.5%; Score 108.4; DB 3; Length 41345;  
Best Local Similarity 55.6%; Pred. No. 2.3e-12;  
Matches 208; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
Qy 474 gccgtggagggtgccaaaccccgctcttcacgtctcttccttcggtgagcccgagcgaag 533  
Db 15583 GCACCTTCAATTTCTTACCGTCTCATTCATTTATGTGTCTGCTGATGCCGATCAAG 15642  
Qy 534 gactaccacaagaagaatgaactcagctcagctcagctcagctcagctcagctcagctcagctc 593  
Db 15643 GATTACCACTAGAGATGAAGTCTTAAATACAAATTAAGAGAACAAAGCTCAGTTGTT 15702  
Qy 594 ttcgtgactgagtgaggagcgtgagcagccagccagccagccagccagccagccagccagccag 653  
Db 15703 TTCGTGATGACTGGTGGTAAAGAAACACATCCAGGATCCGATACATATGAAAAA 15762  
Qy 654 atcgctccacagcttgcgcaagcttcagctcagctcagctcagctcagctcagctcagctcag 713  
Db 15763 ATCGCGCTGCTATCTTTTGGCAAGATTTTCCATCTTGAAGAGAGTGTGATGATACCGTA 15822  
Qy 714 ttaaagtgggtgagtcgagtcagcagccagccagccagccagccagccagccagccagccag 773  
Db 15823 TTAGAAATATGTCAGACATGCTGTCAACACAGAGAAAGTTCATTTGATGTACGAGCAGT 15882  
Qy 774 gagggaggggggaacacacatggagaatcccttttgacccagcttgaaagaagtcacc 833  
Db 15883 GAACGGGAGGAACAGTGTCTAGAAACATCCAGTAGAACAGCATCTCTCAGAGCTCACC 15942  
Qy 834 atctcaatgagcgg 847  
Db 15943 ATCTCATTTGTCAGG 15956

RESULT 9  
LOCUS HSA245418  
DEFINITION Homo sapiens mRNA for G7c protein (G7c gene located in the class  
III region of the major histocompatibility complex).  
ACCESSION AJ245418  
VERSION AJ245418.1 GI:5701855  
KEYWORDS G7c gene; G7c protein.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1610)  
AUTHORS Albertella,M.R., Jones,H., Thomson,W., Olavesen,M.G. and  
Campbell,R.D.  
TITLE Localization of eight additional genes in the human major

histocompatibility complex, including the gene encoding the casein  
kinase II beta subunit (CSNK2B)  
Genomics 36 (2), 240-251 (1996)  
95411681  
MEDLINE 2 (bases 1 to 1610)  
REFERENCE Albertella,M.R.  
JOURNAL Thesis (1997) University of Oxford, Department of Biochemistry, MRC  
Immunochimistry Unit  
3 (bases 1 to 1610)  
AUTHORS Snoek,M., Albertella,M.R., van Kooij,M., Wixon,J., van Vugt,H., de  
Groot,K. and Campbell,R.D.  
TITLE G7c, a novel gene in the mouse and human major histocompatibility  
complex class III region, possibly controlling lung tumor  
susceptibility  
Immunogenetics 51 (4-5), 383-386 (2000)  
JOURNAL 20260998  
MEDLINE 4 (bases 1 to 1610)  
REFERENCE Aquado,B.  
AUTHORS Direct Submission  
TITLE Submitted (02-AUG-1999) Aquado B., HGMP Resource Centre, MRC,  
JOURNAL Genome Campus, Hinxton, Cambridge, CB10 1SB, UNITED KINGDOM  
FEATURES  
source  
1. .1610  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
/note="Class III region of the major histocompatibility  
complex"  
1. .1610  
/gene="G7c"  
<1. .>1610  
/gene="G7c"  
/function="unknown"  
/codon\_start=-3  
/product="G7c protein"  
/protein\_id="CA52192.1"  
/db\_xref="GI:5701856"  
/translation="LGRDPSRLDITPASSLSFVLDITGSMGEINAAKIQARHLVE  
ORGSMPVHYVLVPFHPDGFPVTTSDPSFQOLNEIHALGGDEPEMCLSAQ  
LALLHTPLSDIFVFTDASPKDAFLTNOVSLTOERRCRVTLVTEDTSRVQGRARE  
ILSLRFEPIKAVALSAGSEVIFTKDQHIRDVAALVIGESMAALVTLPLDPVYVPGP  
LVTSVDLLQKITVRIHGDISSFWKNPAGVSGQGRKSSSHTRFGQFPMWMTDDP  
PQTGWEIQVTAEDTPGVVQAQSLDLFLHFGIPMEDGPHGPIPLTQPVAGLTQL  
LVEVTGLSRANPGDPQHFHSHVILRGVPEAGELQVLEVPPEPGLLAASLPTL  
LSTPRPFLSLIGODAAAGRLHRAAPQSTVPVLLLESGPSGLAPGSKVPLSLRIA  
SFSGPDLRLTFTVNPFSLSLTSLRAHLENECLGPAVAGGPRFSFGPRGDDGDCD  
RGRSPSTPDSCLFLLVSPQPRGTG"  
BASE COUNT 290 a 525 c 463 g 332 t  
ORIGIN  
Query Match 9.1%; Score 93.6; DB 9; Length 1610;  
Best Local Similarity 52.8%; Pred. No. 5.2e-09;  
Matches 226; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
Qy 183 acgctgtcccccgcagagagagcgcacccctgcttgccttcgtatgtaccggtcc 242  
Db 24 AGGCTGCTGGACATCACCCAGCCTCCAGCTGAGCTTGTCTCGACACACCGGCAGC 83  
Qy 243 atgtggagcatctgatcgagtgatcgagcgcgctcagcattcttgagcgcagctcg 302  
Db 84 ATGGGTGAGGAGATCAACGCTGCCAAAATCCAGGTCGCCACCTTGTGGAGC---AGCG 140  
Qy 303 agcagccagcgggtcatcgcacacatcgctgggtgccttcctccagcaccagacatt 362  
Db 141 AGAGCAGCCCCATGAGCCTGTCACATATGCTTGGTGGCCCTTTTCATGACCCAGGGTTC 200  
Qy 363 ggccttgagcctcagcgcgagccagtggtgtttcagagagagctgagagcaactctat 422  
Db 201 GGCCTGTCTTTACACACAGTCACCCCTGCAGCTTCTGCGACAGCTTAATGAGATCCAT 260  
Qy 423 gtccagggagggtggtgactgccccaaaaatagtgctggtgggccccataaagcgtccgtg 482





## JOURNAL

Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

On May 4, 2000 this sequence version replaced gi:7634143. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>

GSI-164L12 is from the library Genome\_Systems\_ReleaseI VECTOR: phelobAC11

This sequence is the entire insert of clone GSI-164L12 The true left end of clone GSI-15303 is at 103097 in this sequence.

## FEATURES

## source

Location/Qualifiers  
1..157029  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="GSI-164L12"

## repeat\_region

1..798  
/note="L1p repeat: matches 1..791 of consensus"

## repeat\_region

1064..1512  
/note="LINE2 repeat: matches 5602..6150 of consensus"

## misc\_feature

Complement(2521..2710)

## misc\_feature

2792..3250  
/note="match: GSS: Em:AQ761942"

## repeat\_region

3434..3489  
/note="match: GSS: Em:AQ246643"

## repeat\_region

3435..3488  
/note="14 copies 4 mer tgtg 87% conserved"

## repeat\_region

3435..3488  
/note="27 copies 2 mer tg 88% conserved"

## repeat\_region

4119..4244  
/note="MIR repeat: matches 23..152 of consensus"

## repeat\_region

6902..7098  
/note="MER20 repeat: matches 8..218 of consensus"

## repeat\_region

7164..7191  
/note="7 copies 4 mer ttat 96% conserved"

## repeat\_region

7500..7576  
/note="L2 repeat: matches 2645..2739 of consensus"

## repeat\_region

7735..8041  
/note="AluSq repeat: matches 1..308 of consensus"

## repeat\_region

9440..9901  
/note="MLR1B repeat: matches 1..505 of consensus"

## repeat\_region

10595..10771  
/note="MLR2 repeat: matches 277..553 of consensus"

## repeat\_region

10808..11111  
/note="MLR2 repeat: matches 1..289 of consensus"

## repeat\_region

12089..12200  
/note="L2 repeat: matches 2607..2749 of consensus"

## repeat\_region

12255..12361  
/note="L1pB3 repeat: matches 6044..6150 of consensus"

## repeat\_region

12362..12675  
/note="MER1B repeat: matches 1..337 of consensus"

## repeat\_region

12676..12916  
/note="L1pB3 repeat: matches 5804..6044 of consensus"

## repeat\_region

13006..13612  
/note="L1pB3 repeat: matches 5804..6044 of consensus"

## repeat\_region

/note="L1M1 repeat: matches 5180..5793 of consensus"

## repeat\_region

13614..13877  
/note="AluJo repeat: matches 27..287 of consensus"

## repeat\_region

13879..13981  
/note="L1M2 repeat: matches 2048..2147 of consensus"

## repeat\_region

13982..14292  
/note="AluYa5 repeat: matches 1..311 of consensus"

## repeat\_region

14293..15642  
/note="L1M2 repeat: matches 640..2048 of consensus"

## repeat\_region

15608..15818  
/note="L1PA15-16 repeat: matches 146..337 of consensus"

## repeat\_region

15828..16373  
/note="L1PBA repeat: matches -752..-208 of consensus"

## repeat\_region

16464..16704  
/note="L1M2 repeat: matches -676..-428 of consensus"

## misc\_feature

16497..17042  
/note="match: GSS: Em:AQ004976"

## repeat\_region

17721..19291  
/note="L2 repeat: matches 1051..2708 of consensus"

## repeat\_region

19471..19528  
/note="29 copies 2 mer aa 81% conserved"

## misc\_feature

Complement(21071..21467)

## repeat\_region

21985..22024  
/note="match: GSS: Em:AQ761948"

## repeat\_region

21985..22024  
/note="20 copies 2 mer ct 77% conserved"

## repeat\_region

22384..22611  
/note="MIR repeat: matches 3..262 of consensus"

## repeat\_region

22697..22863  
/note="AluJb repeat: matches 137..302 of consensus"

## repeat\_region

23754..23861  
/note="27 copies 4 mer acat 61% conserved"

## repeat\_region

24319..24381  
/note="AluJb repeat: matches 84..148 of consensus"

## repeat\_region

24382..24670  
/note="AluSx repeat: matches 22..309 of consensus"

## repeat\_region

24671..24822  
/note="AluJb repeat: matches 148..289 of consensus"

## repeat\_region

26950..26985  
/note="18 copies 2 mer tg 97% conserved"

## repeat\_region

26952..26987  
/note="9 copies 4 mer tgtg 97% conserved"

## repeat\_region

27687..27917  
/note="AluJo repeat: matches 83..304 of consensus"

## repeat\_region

29345..29476  
/note="L2 repeat: matches 2605..2750 of consensus"

## misc\_feature

30393..30828  
/note="match: GSS: Em:AQ825928"

## repeat\_region

30990..31275  
/note="AluSx repeat: matches 14..299 of consensus"

## misc\_feature

32321..32866  
/note="match: GSS: Em:AQ418559"

## repeat\_region

33201..33344  
/note="L2 repeat: matches 2562..2696 of consensus"

## repeat\_region

33363..33795  
/note="L2 repeat: matches 2250..2690 of consensus"

## repeat\_region

35271..35548  
/note="AluSq repeat: matches 18..297 of consensus"

## misc\_feature

Complement(37223..37713)

## misc\_feature

37319..37339  
/note="match: GSS: Em:AQ667379"

## repeat\_region

38212..38679  
/note="match: STS: Em:G37800"

## repeat\_region

40470..40593  
/note="L1M1 repeat: matches 5847..6324 of consensus"

## repeat\_region

41014..41123  
/note="MIR repeat: matches 8..140 of consensus"

## repeat\_region

41395..41618  
/note="MIR repeat: matches 20..141 of consensus"

## repeat\_region

41395..41618  
/note="MIR repeat: matches 10..256 of consensus"

## repeat\_region

43230..43494  
/note="AluY repeat: matches 39..307 of consensus"

## repeat\_region

43683..44142  
/note="L1MEC repeat: matches 1211..1677 of consensus"

```

repeat_region 44767..45295
/note="L1 repeat: matches 2530. .3129 of consensus"
repeat_region 45333..45571
/note="AluJb repeat: matches 3. .41 of consensus"
repeat_region 45604..46929
/note="L1PA7 repeat: matches 4758. .6124 of consensus"
repeat_region 46930..47199
/note="AluJb repeat: matches 24. .297 of consensus"
repeat_region 47356..47823
/note="L1 repeat: matches 3647. .4127 of consensus"
repeat_region 47918..48310
/note="L1 repeat: matches 4257. .4654 of consensus"
repeat_region 48311..48614
/note="AluX repeat: matches 1. .308 of consensus"
repeat_region 48615..48796
/note="L1 repeat: matches 4654. .4834 of consensus"
misc_feature 49029..49516
/note="match: GSS: Em:AQ280129"
repeat_region 49087..49262
/note="L1P repeat: matches 5164. .5357 of consensus"
repeat_region 49407..49510
/note="L2 repeat: matches 2579. .2696 of consensus"
repeat_region 50481..51060
/note="MLTIF repeat: matches 1. .536 of consensus"
misc_feature complement(51651..52071)
/note="match: GSS: Em:AQ636105"
misc_feature 52290..52512
/note="match: GSS: Em:AQ797420"
repeat_region 52936..53296
/note="L1PA15 repeat: matches 5785. .6156 of consensus"
repeat_region 53303..53330
/note="7 copies 4 mer gtgt 96% conserved"
repeat_region 53355..53457
/note="L1PA15 repeat: matches 5689. .5791 of consensus"
repeat_region 53680..53695
/note="108 copies 2 mer tt 57% conserved"
misc_feature complement(54152..54622)
/note="match: GSS: Em:AQ568896"
repeat_region 54157..54274
/note="MER45B repeat: matches 199. .318 of consensus"
repeat_region 54340..54856

```

Query Match 8.7%; Score 89.4; DB 9; Length 157029;  
Best Local Similarity 64.0%; Pred. No.1e-08;  
Matches 135; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

Qy 709 aggtgttaaagtgggtgagtcgccatccaggtcccaagttcattctgtctgcagcag 768
Db 125632 AGGTATTAAATGGGTAGGAAGCAGTACAGGCTCCAAAGTTCACTTTTATCCACAG 125691

Qy 769 accacgaggaggcggaacacacatggagatcccttttgacccagcttgaggag 828
Db 125692 ATCATTTGGAACAGGCTGAATACTTGGAGAAATCTTTTGATCCCGCTGAAAGAGG 125751

Qy 829 tcaccatctactgagcgccagggcctgagatcgatcgagtcgagtcggagccactgggtatgt 888
Db 125752 TCACGTGTGCTTTGAGTGGGCTCTCCAAATGATTGAATTCGCAATCTTTAGTGCAGA 125811

Qy 889 cccagggttcacctctctctctgtgcaaga 919
Db 125812 TATATCAACATCATCAATAATAATACAAA 125842

```

RESULT 12  
HSG118G19  
LOCUS  
DEFINITION HSG118G19 87461 bp DNA PRI 15-MAY-2001  
Human DNA sequence from clone GSI-118G19 on chromosome 1q25.1-31.1,  
complete sequence.  
ACCESSION AL121996  
VERSION AL121996.15 GI:14140106  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 87461)  
Thomas,D.  
Direct Submission  
Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On May 17, 2001 this sequence version replaced gi:13990347.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one W13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
GSI-118G19 is from the library Genome\_Systems\_ReleaseI  
pBelOBAC11  
IMPORTANT: This sequence is not the entire insert of clone  
GSI-118G19. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone GSI-16412 is at 87362 in this sequence.  
The true right end of clone GSI-204112 is at 100 in this sequence.  
Location/Qualifiers  
1..87461  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q25.1-31.1"  
/clone="GSI-118G19"  
/clone\_lib="Genome\_Systems\_ReleaseI"  
5..571  
/note="L2 repeat: matches 2120. .2748 of consensus"  
568..694  
/note="L2 repeat: matches 2572. .2706 of consensus"  
896..1200  
/note="AluJb repeat: matches 1. .300 of consensus"  
1292..1578  
/note="AluJb repeat: matches 19. .297 of consensus"  
1618..2109  
/note="L1MA9 repeat: matches 5783. .6270 of consensus"  
2110..2157  
/note="24 copies 2 mer gt 95% conserved"  
2166..2326  
/note="AluSg/x repeat: matches 149. .309 of consensus"  
2329..2936  
/note="L1MA9 repeat: matches 5177. .5779 of consensus"  
2936..3566  
/note="L1MA9 repeat: matches 4500. .5146 of consensus"  
3794..3891  
/note="MER5A repeat: matches 13. .112 of consensus"  
5634..5814  
/note="MIR repeat: matches 59. .247 of consensus"  
complement(5905..5949)  
/note="Sequence confirmed by AC023275 sequenced by WUGSC"  
6728..6926  
/note="MIR repeat: matches 56. .259 of consensus"  
7074..7136  
repeat\_region

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

```
repeat_region /note="MIR repeat: matches 50. .112 of consensus"
8362. .8848
repeat_region /note="LTR37B repeat: matches 6. .459 of consensus"
10408. .10856
repeat_region /note="MLT1C repeat: matches 2. .466 of consensus"
10884. .11020
repeat_region /note="MLT2B repeat: matches 210. .363 of consensus"
11056. .13925
repeat_region /note="L1PA3 repeat: matches 13. .2882 of consensus"
13926. .13962
repeat_region /note="Charlie3 repeat: matches 878. .914 of consensus"
13963. .17229
misc_feature /note="L1PA3 repeat: matches 2881. .6146 of consensus"
17162. .17277
misc_feature /note="Sequence confirmed by AC023275 sequenced by WUGSC"
17278. .17310
misc_feature /note="Sequence from AC023275 sequenced by WUGSC."
17311. .17418
repeat_region /note="Sequence confirmed by AC023275 sequenced by WUGSC"
17930. .18002
repeat_region /note="L1P4 repeat: matches 5695. .5768 of consensus"
18310. .18794
repeat_region /note="MLT1H repeat: matches 51. .543 of consensus"
20658. .20691
repeat_region /note="L7 copies 2 mer tg 94% conserved"
20801. .21109
repeat_region /note="AluYb repeat: matches 1. .312 of consensus"
23465. .23676
repeat_region /note="MIR repeat: matches 2. .258 of consensus"
23750. .23852
repeat_region /note="L2 repeat: matches 2596. .2710 of consensus"
26376. .26824
repeat_region /note="MER65A repeat: matches 1. .445 of consensus"
26984. .27214
repeat_region /note="L1MC3 repeat: matches 7496. .7733 of consensus"
27243. .27862
repeat_region /note="L1MC3 repeat: matches 6695. .7336 of consensus"
27895. .27991
repeat_region /note="L1MC4 repeat: matches 6613. .6709 of consensus"
28061. .28121
repeat_region /note="L1MD3 repeat: matches 6536. .6597 of consensus"
29391. .29470
repeat_region /note="FLAM4 repeat: matches 42. .121 of consensus"
29472. .29787
repeat_region /note="L58 copies 2 mer at 78% conserved"
30307. .30453
repeat_region /note="MIR repeat: matches 79. .213 of consensus"
30455. .30751
repeat_region /note="AluX repeat: matches 1. .297 of consensus"
30775. .30836
repeat_region /note="MIR repeat: matches 196. .260 of consensus"
30908. .30959
repeat_region /note="MIR repeat: matches 82. .140 of consensus"
32816. .32938
repeat_region /note="MIR repeat: matches 17. .142 of consensus"
33149. .33494
repeat_region /note="L2 repeat: matches 2355. .2710 of consensus"
34283. .35033
repeat_region /note="L1MEC repeat: matches 783. .1570 of consensus"
35069. .36448
repeat_region /note="L1MEC repeat: matches 1635. .2727 of consensus"
36454. .36607
repeat_region /note="MIR repeat: matches 8. .171 of consensus"
40801. .41093
repeat_region /note="AluYb repeat: matches 15. .312 of consensus"
41795. .42230
repeat_region /note="L1MB8 repeat: matches 5745. .6173 of consensus"
42231. .42396
repeat_region /note="L1M4 repeat: matches 5483. .5646 of consensus"
42531. .42960
repeat_region /note="L1PA16 repeat: matches 5721. .6157 of consensus"
43280. .44137
repeat_region /note="L2 repeat: matches 1880. .2750 of consensus"
```

```
repeat_region 44142. .44298
/note="MSTD repeat: matches 1. .134 of consensus"
repeat_region 44374. .44680
/note="AluX repeat: matches 5. .311 of consensus"
repeat_region 44684. .44934
/note="MSTD repeat: matches 178. .426 of consensus"
repeat_region 44935. .45514
/note="L2 repeat: matches 1247. .1880 of consensus"
repeat_region 46831. .46872
/note="L2 repeat: matches 2709. .2750 of consensus"
repeat_region 47636. .47950
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 48164. .48425
/note="L2 repeat: matches 2431. .2710 of consensus"
repeat_region 48443. .48632
/note="L2 repeat: matches 2027. .2233 of consensus"
repeat_region 49480. .49851
/note="L2 repeat: matches 1235. .1662 of consensus"
repeat_region 49871. .49964
/note="MIR repeat: matches 20. .127 of consensus"
repeat_region 49991. .50252
/note="MLT1C repeat: matches 173. .456 of consensus"
repeat_region 50253. .50564
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 50565. .50743
/note="MLT1C repeat: matches 1. .173 of consensus"
repeat_region 50981. .51277
/note="AluX repeat: matches 1. .296 of consensus"
repeat_region 51693. .51718
/note="L3 copies 2 mer tt 100% conserved"
repeat_region 52326. .52507
/note="MIR repeat: matches 76. .244 of consensus"
repeat_region 52558. .52629
/note="MIR repeat: matches 75. .149 of consensus"
repeat_region 52640. .52807
/note="MER63A repeat: matches 6. .174 of consensus"
repeat_region 53094. .53380
/note="L2 repeat: matches 1729. .2035 of consensus"
repeat_region 53456. .53885
/note="L2 repeat: matches 2285. .2730 of consensus"
repeat_region 54316. .54481
/note="83 copies 2 mer tt 56% conserved"
repeat_region 55309. .55565
/note="L1MD repeat: matches 1211. .1474 of consensus"
repeat_region 56353. .56911
/note="L1M4 repeat: matches 3491. .4067 of consensus"

Query Match 6.8%; Score 69.8; DB 9; Length 87461;
Best Local Similarity 63.3%; Pred. No. 0.0001;
Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 196 ccacagagacgcacacctgcttgcgtatgtcaccggtccatgtggagcgc 255
|| || || || || || || || || || || || || || || || || || || ||
Db 38544 CCGAGGGGGCTCCACCTTGGCTTTGTTGATGTGACTGTCTTATGATGATT 38603

QY 256 tgatgcaggtgacgcgcgcctcgcagcagcagcagcagcagcagcagcc 315
|| || || || || || || || || || || || || || || || || || || ||
Db 38604 TAGTTCAGGTGATGAGGGGGCTCCAAAATTTGGAGACGCTCTTGAAGACCTATAA 38663

QY 316 gggatcgcgcacactgcgtgctgttccacgaccagacattgg 364
|| || || || || || || || || || || || || || || || || || || ||
Db 38664 GACTCTTTTCAACTTTGGCTTGGTGGCTTTCCATGATCCAGGTAAAGG 38712

RESULT 13
AC023275/c
LOCUS AC023275 180248 bp DNA PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-375110 from 1, complete sequence.
ACCESSION AC023275
VERSION AC023275.3 GI:9838169
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 180248)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

2 (bases 1 to 180248)

Swearengen, S., Mauplin, R., Drone, K. and Gregory, S.

The sequence of Homo sapiens BAC clone RP11-375L10

Unpublished

3 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (10-FEB-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (17-AUG-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (10-SEP-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

6 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (08-NOV-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

7 (bases 1 to 180248)

Waterston, R.H.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Aug 17, 2000 this sequence version replaced gi:7630968.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

----- Summary Statistics

----- Center project name: H\_NH0375L10

-----

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this

sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from

Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong

and coworkers at the Roswell Park Cancer Institute

(<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-375L10;

Actual end is at base position 180248 of RP11-375L10.

Location/Qualifiers

1. 180248

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1"

/clone="RP11-375L10"

/clone\_lib="RPCI-11"

667. 820

/rpt\_family="CRI"

926. 1011

/rpt\_family="CRI"

2272. 2420

/rpt\_family="MIR"

5140. 5262

/rpt\_family="L2"

6076. 6593

/rpt\_family="L2"

6767. 7258

/rpt\_family="L2"

7348. 7653

/rpt\_family="Alu"

7848. 8645

/rpt\_family="L2"

8938. 9011

/rpt\_family="MER1\_type"

9685. 9725

/rpt\_family="L1"

9801. 10110

/rpt\_family="Alu"

12246. 12330

/rpt\_family="MIR"

12577. 12788

/rpt\_family="Alu"

12790. 12957

/rpt\_family="Alu"

13055. 13209

/rpt\_family="L2"

16393. 17278

/rpt\_family="Achobo"

17344. 17461

/rpt\_family="Achobo"

18161. 18263

/rpt\_family="L1"

18322. 18687

/rpt\_family="L1"

20558. 21137

/rpt\_family="MaLR"

22114. 22321

/rpt\_family="L2"

22353. 22811

/rpt\_family="L1"

22832. 23003

/rpt\_family="L1"

23004. 23307

/rpt\_family="Alu"

23308. 23700

/rpt\_family="L1"

23740. 24219

/rpt\_family="L1"

24419. 24688

/rpt\_family="Alu"

24689. 26014

/rpt\_family="L1"

26047. 26106

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```
/rpt_family="Alu"
26433..26642
/rpt_family="L1"
26654..26881
/rpt_family="L1"
27044..27199
/rpt_family="L1"
27270..27726
/rpt_family="L1"
28124..28388
/rpt_family="Alu"
29167..29275
/rpt_family="MIR"
30000..30223
/rpt_family="MIR"
30495..30677
/rpt_family="MIR"
31025..31217
/rpt_family="MIR"
32939..33406
/rpt_family="L1"
33434..33836
/notes="similar to
34177..34516
/notes="similar to
35831..35967
/rpt_family="CR1"
36070..36347
/rpt_family="Alu"
36643..36920
/rpt_family="CR1"
36969..37038
/rpt_family="CR1"
37823..38255
/rpt_family="L2"
38274..38387
/rpt_family="L2"
39802..39869
/rpt_family="L2"
40343..40628
/rpt_family="Alu"
43701..43931
/rpt_family="Alu"
46796..46947
/rpt_family="Alu"
46948..47236
/rpt_family="Alu"

Query Match      6.8%; Score 69.8; DB 9; Length 180248;
Best Local Similarity 63.3%; Pred. No. 8.3e-05;
Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 196 ccacaggagagccaccctggccttcgatgtcacgcggtcccatgtggacgac 255
|| || || || || || || || || || || || || || || || || || || ||
Db 120435 CCGAGGGGGCTCCACGCTGGCTTTGGTTGATGTGACGTGCTCTATGATGATT 120376

Qy 256 tgatcgagtgatgacgcgccttcacgcattctggacgcagcttgagcagcgaccc 315
|| || || || || || || || || || || || || || || || || || || ||
Db 120375 TAGTTCAGTGATGAGAGGGGCTCCAAAATTTGGAGACGCTCTTGAAGACCTAAAA 120316

Qy 316 gggtcacgcacactatcgctgggtgcctttccacgaccagacattgg 364
|| || || || || || || || || || || || || || || || || || || ||
Db 120315 GACCTCTTTCACTTTCGGTTGGTGCTTCCATGATCCAGGTAAGGG 120267

RESULT 14
LOCUS HSJ34G10 717 bp DNA STS
DEFINITION STS from H.sapiens random shear fragment, sequence tagged site.
ACCESSION AL159308
VERSION AL159308.1 GI:7210622
KEYWORDS STS.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 717)
AUTHORS Hunt,S., Sims,S., Willey,D., Carter,N. and Ross,M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2000) E-mail contact: humquery@sanger.ac.uk
COMMENT Marker stSG68675 (Primer A : CTCCTCCAGATGGAGCC;
Primer B : TGGATGTTATATCGTCTCG; amplicon size : 121 bp) was
developed from a single pass sequencing read from H.sapiens
flow-sorted chromosome 9-12 random shear fragment, SC9-12pJ34g10.
Vector : pUC18 Site : SmaI
Further information : http://www.sanger.ac.uk/HGP/Chr9/.
FEATURES
Source 1..717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/sex="male"
/clone="SC9-12pJ34g10"
/tissue_type="Lymphoblastoid cell line"
/clone_lib="SC9-12pJ"
BASE COUNT 186 a 198 c 211 g 120 t 2 others
ORIGIN
Query Match 6.3%; Score 65.4; DB 11; Length 717;
Best Local Similarity 72.4%; Pred. No. 0.0028;
Matches 84; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 330 tatcgctgggtcccttcacgaccagacattgcccagtcacccctcacgcgcgaccca 389
|| || || || || || || || || || || || || || || || || || || ||
Db 557 TCTGGGGCTCCCTCTCTCTCCACAGATATTGGCNCAGTGACCTCAGCGGACCC 498

Qy 390 gtggtgttcagagagctgagacacactatgttcaggaggtggagctgacctgcc 445
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 ACAGTGTTCAGAGGAGCTGAGAGAACTCTACGTGCAGGTGGCGAGCCCTGACC 442

-RESULT 15
LOCUS HS348TG1 356 bp DNA STS
DEFINITION H.sapiens (DIS2848) DNA segment containing (CA) repeat; clone
AFM348tg1; single read, sequence tagged site.
ACCESSION Z51502
VERSION Z51502.1 GI:1232802
KEYWORDS STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
microsatellite marker; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 356)
AUTHORS Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
E-mail: Jean.Weissenbach@genethon.fr
Dbb,C., Faure,S., Fizames,C., Samson,D., Drouot,N., Vignal,A.,
Mollasseau,P., Marc,S., Hazan,J., Seboun,E., Lathrop,M., Gyapay,G.,
Morissette,J. and Weissenbach,J.
A comprehensive genetic map of the human genome based on 5,264
microsatellites
Nature 380 (6570), 152-154 (1996)
96176476
full automatic.
Location/Qualifiers
Source 1..356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/cell_line="CEPH 134702"
/clone_lib="genomic DNA"
```

BASE COUNT 111 a /note="cloning vector is M13mpl8"  
 ORIGIN 79 c 46 g 107 t 13 others

Query Match 6.0%; Score 61.6; DB 11; Length 356;  
 Best Local Similarity 68.7%; Pred.No. 0.02; Mismatches 36; Indels 0; Gaps 0;  
 Matches 79; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 580 agtcgagggtggtcttcgtgactgggactgagcgacccacccctggctacc 639  
 Db 242 AATCATAGGTCCTATTGTTCTGCTGNGATGATGACAGNCCCATATNGGTATA 301  
 QY 640 tggttttgaggagatgcctccaccagttctggcccaagtgtccagctggacaa 694  
 Db 302 AAGTCTATGGNGAAATGCCTCTNCAAGTTCTGTCAAGTGTTCATCTGGACAA 356

Search completed: March 7, 2002, 20:44:19  
 Job time: 6992 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 18:53:02 ; Search time 1634.39 Seconds  
(without alignments)  
6778.613 Million cell updates/sec

Title: US-09-665-728-2

Perfect score: 1031

Sequence: 1 tctagcgacccttcggcc.....agccggctcgagcgccgc 1031

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_estl:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	309	30.0	470	AI449932	AI449932 mr81b03.x
C 2	90.4	8.8	720	AZ985073	AZ985073 2M0266L17
C 3	77.4	7.5	937	CNS025DE	AL181931 Tetraodon
C 4	70.6	6.8	559	BF418902	BF418902 UI-R-BJ2-
C 5	64	6.2	915	AZ137837	AZ137837 SP-0178-A
C 6	61	5.9	528	BE508854	BE508854 dcl3h07.y
C 7	55	5.3	754	AZ189426	AZ189426 SP-1014_A
C 8	49.6	4.8	935	CNS006AK	AL066051 Drosophil
C 9	48.8	4.7	925	CNS0091P	AL053013 Drosophil
C 10	46.6	4.5	520	BG895162	BG895162 358378 MA
C 11	46.4	4.5	925	CNS0091P	AL053013 Drosophil
C 12	46	4.5	965	CNS02215	AL178322 Tetraodon

C 13	45	4.4	818	13	CNS03902	AL234371	Tetraodon
C 14	44.8	4.3	375	11	BF481079	BF481079	FM1_16_B0
C 15	44.8	4.3	379	11	BG560654	PHIZ2_58	BG560654
C 16	44.8	4.3	392	11	BF481089	FM1_16_CO	BF481089
C 17	44.8	4.3	406	11	BG560698	RHIZ2_58	BG560698
C 18	43.2	4.2	523	11	BG240204	OVI_19_B0	BG240204
C 19	43.2	4.2	924	13	CNS04SRD	Tetraodon	AL124542
C 20	42.6	4.1	543	10	AW124542	UI-M-BH2	AW124542
C 21	42.2	4.1	461	10	AU174537	AU174537	AU174537
C 22	42	4.1	632	11	BG744240	602723347	BG744240
C 23	41.8	4.1	503	10	AI653725	WD36F06.x	AI653725
C 24	41.8	4.1	686	11	BG835990	2m06_06h0	BG835990
C 25	41.4	4.0	384	10	AA294693	SMOV3MCA9	AA294693
C 26	41.4	4.0	476	10	BE477180	160761 BA	BE477180
C 27	41.4	4.0	566	10	AW082912	xc04c11.x	AW082912
C 28	41.4	4.0	872	11	BG345072	HVSM8G001	BG345072
C 29	41	4.0	647	11	BG180123	602329707	BG180123
C 30	41	4.0	778	10	AI991270	wu41h01.x	AI991270
C 31	41	4.0	814	10	AW081400	xc41h07.x	AW081400
C 32	40.8	4.0	523	10	AI855212	603010C06	AI855212
C 33	40.8	4.0	556	11	BG874151	MEST46-H0	BG874151
C 34	40.8	4.0	584	13	AQ848793	LMAJFV1_1	AQ848793
C 35	40.8	4.0	586	10	BE553066	946088H08	BE553066
C 36	40.8	4.0	603	10	AI855218	603010D06	AI855218
C 37	40.8	4.0	626	10	AI855245	603010H10	AI855245
C 38	40.8	4.0	671	13	AQ901789	LMAJFV1_1	AQ901789
C 39	40.8	4.0	813	11	BG321263	2m04_0390	BG321263
C 40	40.8	4.0	956	13	CNS03GMI	Tetraodon	AL243171
C 41	40.6	3.9	460	10	AI432376	tg54h07.x	AI432376
C 42	40.6	3.9	573	10	BE452815	894066A05	BE452815
C 43	40.6	3.9	738	11	BG864336	602798212	BG864336
C 44	40.4	3.9	459	10	AW504911	UI-HF-BN0	AW504911
C 45	40.4	3.9	470	11	BF313055	601896455	BF313055

#### ALIGNMENTS

RESULT 1  
AI449932/c  
LOCUS  
DEFINITION  
AI449932 470 bp mRNA  
IMAGE:603821.3', similar to WP:FI5G9.4 CE01552 IG SUPERFAMILY  
REPEATS: // mRNA sequence.  
AI449932.1 GI:4293462  
EST.  
house musculus  
house musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 470)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,  
E., Kohn, S., Shin, T., Jackson, K., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Possible reversed clone: polyT not found  
High quality sequence stop: 467.  
Location/Qualifiers  
1..470

```
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:603821"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
EcoRI; Site_2; XhoI; Vector: phuscript SK-; Site_1:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5',
adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'
BASE COUNT      80 a 152 c 168 g 69 t 1 others
ORIGIN
Query Match      30.0%; Score 309; DB 10; Length 470;
Best Local Similarity 92.8%; Pred. No. 2.9e-57;
Matches 324; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
Qy 19 ccgctagagagactgcactgcccattatccctgcgacctgcgctcccaattaggcgt 78
Db 386 CCCGTAGAGCGAGACTGCACCTGCCATCTCTCACTGCACCTCGGGTCCCATCAGGGCT 327
Qy 79 gcagcctcgctcagatgacgcttagggcgagctcctgcgctgctcctgcggaact 138
Db 326 GCAGCCTCCGGCTACCATGAGCGCTGGGGCGCAGCTCCTGCGCTGCTGCGGATCT 267
Qy 139 acacagttagggcgggcggtccacatctgtgagccacgacgagcgtgtcccccgcga 198
Db 266 CTACAGCAGTGGCGGGTGTGACGTCTGACGCGCCGCCACGAGACGCTGTCCCGGCCA 207
Qy 199 caggagagccacccctggcctctgtcttgatgtcacgggtccatgtggagcagctga 258
Db 206 CAGGAGAGCCACCCCTAGCCTTCGTTCGATGTCAACGGGTCCATGTGGGATGACCTGA 147
Qy 259 tgcagtgatgagcgccctcagcattctggagcagctgagcagcgcgcgcgcggg 318
Db 146 TGCAGGTGATGACGGCGCCCTCGCGCATCTGGAGCGAGCTGAGACGCCGACGCCGG 87
Qy 319 tcatgcccaactatgcgtggtgctcttccacgacccagacattggccc 367
Db 86 TCATGCCCAACTATGCGTGTGCTGCTTTCACGAGCCACCCAGGTAGAGCCCC 38
RESULT 2
AZ985073/c      720 bp      DNA      GSS      27-APR-2001
LOCUS      2M0266L17R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION      clone UUGC2M0266L17 R, DNA sequence.
ACCESSION      AZ985073
VERSION      AZ985073.1 GI:13856300
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 720)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weis,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
```

```
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0266 row: L column: 17
Seq primer: CACACAGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 720.
Location/Qualifiers
1..720
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0266L17"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, P-"
/notes="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      203 a 134 c 145 g 238 t
ORIGIN
Query Match      8.8%; Score 90.4; DB 13; Length 720;
Best Local Similarity 72.0%; Pred. No. 8.9e-10;
Matches 118; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 426 caggagagtggtgactgccagaaaatgagtgtgggggcccatacaggtgcggtgagggct 485
Db 274 CAGGAGGTGTTGACTGTCCAGAAATGAGCATCGGGGCTATAAAGATTGCCTTGGAAATC 215
Qy 486 gccaaccccggtccttcacatcagctcttcgagatgccctgccaaggaactaccacaag 545
Db 214 TCTCTTCTGGGTCTTCATCTATGTTTTCACGTATGATGACCATCCAGGATTTATCGGCTC 155
Qy 546 aagaatgagctcctgcagctcctgcagctgaagcagctgcaggt 589
Db 154 ACCCATGAGGTGCTGCAGCTTATCCACAGAGAACAAATCTCAGGT 111
RESULT 3
CNS025DE      937 bp      DNA      GSS      12-MAY-2000
LOCUS      CNS025DE
DEFINITION      Tetradon nigroviridis genome survey sequence PUC-Or1 end of clone
236N09 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION      AL181931
VERSION      AL181931.1 GI:7820017
KEYWORDS      GSS; genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM      Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetradon.
1 (bases 1 to 937)
```

## AUTHORS

Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
freshwater pufferfish Tetraodon nigroviridis

## JOURNAL

## REFERENCE

## AUTHORS

2 (bases 1 to 937)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Genoscope.  
3 (bases 1 to 937)  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

## source

Location/Qualifiers  
1..937  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="G"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG236G05SP1-end ;  
PUC-ori"

BASE COUNT 196 a 285 c 260 g 193 t 3 others

## ORIGIN

Query Match 7.5%; Score 77.4; DB 13; Length 937;  
Best Local Similarity 66.5%; Pred. No. 5.9e-07;  
Matches 111; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 191 cccgcacagagagaccacccctgcctcgtctcgtatgcacgcgctccatgtggga 250

Db 453 CCATCCGGAGGCGCATCGACGCTGGCTTCGACGTGACCGCTCCATGTACGA 512

QY 251 cgatcgtatcaggtgatcagcgccctcacgcattcttgagcgcagtcgagcagc 310

Db 513 CGACCTGGTCAGGTCACTCAGGCGCATCCAAAGCTCGAAACATCCCTGAGCAGACC 572

QY 311 cagccgggtcattcgcacattatcgtcgtgctcttcacgacccag 357

Db 573 GAAGAAACCTCTGTATAACTTCGCCCTGGTGCCATTTCATGATCCAG 619

## RESULT 4

## BF418902

## LOCUS

BF418902 559 bp mRNA EST 28-NOV-2000  
UI-R-BJ2-bqk-c-06-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

1 (bases 1 to 559)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Oligo-qt track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com) The following repetitive  
elements were found in this cDNA sequence: 227-278, >(GAAAA

)nSimple.repeat

Seq primer: M13 Forward

POLYA=No.

## FEATURES

## source

Location/Qualifiers  
1..559  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone\_lib="UI-R-BJ2-bqk-c-06-0-UI"  
/lab\_host="UI-R-BJ2"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2  
library is a subtracted library derived from the following  
tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,  
atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15  
dpc, AV canal at 15 dpc. For a detailed description of  
the library from which this clone was derived, please  
visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The  
subtraction has been previously described in (Bonaldo,  
Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_SEQ=None found"

BASE COUNT 143 a 137 c 152 g 127 t

## ORIGIN

Query Match 6.8%; Score 70.6; DB 11; Length 559;  
Best Local Similarity 63.0%; Pred. No. 1.8e-05;  
Matches 109; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 201 ggaagcgcacccctgcctcgtctcgtatgcacgcgctccatgtggagcagctgatg 260

Db 386 GGGCGTCCACGTTGGCTTTTGTGTGAGTGCCTGCCATGATGATGATTTAGTT 445

QY 261 cagtgatcagcgccctcacgcattcttgagcgcagtcgagcagccgagcgggtc 320

Db 446 CAGGTATTGAAGGGCTTCCAAATTTGGAGACGCTTTGAAAGACCTAAGACCT 505

QY 321 atcgcaactatcgctggtgctttccacgacccagacattggccagtgac 373

Db 506 CTTTACAACCTTGTGTTGGTGCCTTCCCATGATCAGAAATCGGTCCAGTGAC 558

## RESULT 5

## AZ137837/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

AZ137837 915 bp DNA GSS 28-AUG-2000  
SP\_0178\_A2\_E10.SP6E Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
genomic clone Plate=178 Col=20 Row=I, DNA sequence.

AZ137837.1 GI:8289740  
GSS.  
Strongylocentrotus purpuratus.  
Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 915)  
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,  
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,  
G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and  
Hood,L.

A sea urchin genome project: Sequence scan, virtual map, and  
additional resources  
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

MEDLINE  
COMMENT

20402566  
Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
Pasadena Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 178 row: I column: 20  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 915.  
Location/Qualifiers  
1..915

FEATURES  
source

/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="plate=178 Col=20 Row=I"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library"  
/note="organ: sperm; vector: BAC3.6; BAC Clones in E-Coli  
DH10B"  
280 a 219 c 174 g 242 t

BASE COUNT  
ORIGIN

Query Match 6.2%; Score 64; DB 13; Length 915;  
Best Local Similarity 62.5%; Pred. No. 0.00048;  
Matches 100; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
Qy 198 acaggagagcgcacctgcttcttcgatgtccacggctccatgtggagacatctg 257  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 457 ACTGCGGAACACGATGCTTTGTTTGTGATGAACGGATCATCTTTGATGACTTG 398  
Qy 258 atgcagtgatgcagcgcctcgcacgcaattcttgagcagctgagcgcgcgcgcg 317  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 397 CAGCAGGTATCGAAGGAGCAGAGCGTATTTTGGATTCATTTGAAGAGAGACACG 338  
Qy 318 gtcacgcgaactatgcctgctggtgctcttcacgaccacg 357  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 337 CCTTTGAAAACCTTTGCTGCTGGTCCCTTTTCATGATCCAG 298

## RESULT 6

BE508854 528 bp mRNA EST 07-AUG-2000  
LOCUS Gc13h07.y1 NICHD XGC Lil Xenopus laevis CDNA clone IMAGE:3397021 5'  
DEFINITION similar to WP:F15G9.4A CE18595 ;, mRNA sequence.  
BE508854  
ACCESSION BE508854.1 GI:9728629  
VERSION EST.  
KEYWORDS African clawed frog.  
SOURCE Xenopus laevis  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 528)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Other ESTs: dc13h07.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available  
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 430.  
Location/Qualifiers  
1..528

FEATURES  
source

/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:3397021"  
/clone\_lib="NICHD XGC Lil"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.4 kb. Constructed by Life  
Technologies. Note: This is a Xenopus Gene Collection (XGC  
) library."  
140 a 109 c 116 g 163 t

BASE COUNT  
ORIGIN

Query Match 5.9%; Score 61; DB 10; Length 528;  
Best Local Similarity 48.7%; Pred. No. 0.0022;  
Matches 166; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
Qy 198 acaggagagcgcacctgcttcttcgatgtccacggctccatgtggagacatctg 257  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 184 ACAGTGACTTCATCTCTGACCTTCTTGTGGACACCATGTTTCCATGGCGATGACCTT 243  
Qy 258 atgcagtgatgcagcgcctcgcacgcaattcttgagcgcagctgagcgcgcgcgcg 317  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 244 CAACAGTTGAAGAGGAGCGCTACACCTGGTTACTTAGCAGCGTTTCTGTTCAGTTCCCATGT 303  
Qy 318 gtcacgcgaactatgcctgctggtgctcttcacgaccacgacattgcccagtgaccctc 377  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 304 GGTGTACGTCACTAGTACAAATGTTGAATTTAATCATCCAGGATTTGGCCCTGCCAGACTT 363  
Qy 378 acg9cggaccacagtggtgttttcagagagagctgagacaaactctatgttcaggaggtggt 437  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 364 ACAAGTTCGGAACCTGATTTAATAATTTTTCNAATCTTAGAGCCTAGATGGTGGT 423  
Qy 438 gactgcccagaaatgagtgtggtggggccatcaaggctgcgtggaggttgccaaaccccgcc 497  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 424 GACTGCCAGAAATATGCAATGGTGGACTTAAGTTGGCTTTCGAGGAATATCACCATAAT 483  
Qy 498 tccttcctacgtctctctcgatgccgtgccaggacta 538  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 484 TCATAATCATGGTCTTACAGATGCTGCCCTTAGGATTA 524

## RESULT 7

AZ189426/c  
LOCUS SP\_1014\_AL\_A06\_SP6E Strongylocentrotus purpuratus, purple sea  
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
genomic clone Plate=1014 Col=11 Row=A, DNA sequence.  
ACCESSION AZ189426  
VERSION AZ189426.1 GI:8372518  
KEYWORDS GSS.  
SOURCE Strongylocentrotus purpuratus.  
ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,  
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray  
,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and  
Hood,L.  
TITLE A sea urchin genome project: Sequence scan, virtual map, and  
additional resources  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
MEDLINE 20402566  
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu





FEATURES	Location/Qualifiers	http://www.genoscope.cns.fr/Tetraodon.
source	1..818	



/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="008E16"

/clone\_lib="G"

BASE COUNT 142 a 104 c 245 g 149 t 178 others

Query Match 4.4%; Score 45; DB 13; Length 818;  
Best Local Similarity 25.9%; Pred. No. 6.4;  
Matches 87; Conservative 101; Mismatches 143; Indels 5; Gaps 1;  
QY 187 tgtccccccacagagagccacccctgcttctgtcttctgtcagtcacccggtccatgt 246  
Db 622 TCTCTCTCAACACTCATYTTTCACCTTCCCTCTCTCATCCCMCCCTTCTCYTTC 563  
QY 247 gggagcatctgacggatgacggcgctcagcattcttgagcagcagcttgagca 306  
Db 562 CCCACAAACATAAASST 503  
QY 307 gcgcagcgggtcgcacactatgcctgctgcttccacacccagacattggcc 366  
Db 502 TSTSSSBTTATTSASSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 443  
QY 367 cagtgacctcagcgcgaccagtggtgttccagagaga-----gctgagacaaactcta 421  
Db 442 ASATSSASASASS 383  
QY 422 tttcagagagctgactgctcagagaaatgagtgggggcatcaaggctgcgctgga 481  
Db 382 ASSNAGASS 323  
QY 482 ggttgccacccggctcctctcatctacgtctctc 517  
Db 322 GSSGCGAGSGSCTCTSACASSCTSSSASSTSTC 287

RESULT 14  
LOCUS BF481079 375 bp mRNA EST 05-DEC-2000  
DEFINITION FM1\_16\_B06.b1\_A003 Floral-Induced Meristem 1 (FMI) Sorghum  
propinquum cDNA, mRNA sequence.  
ACCESSION BF481079  
VERSION BF481079.1 GI:11551900  
KEYWORDS EST.  
SOURCE Sorghum propinquum.  
ORGANISM Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 375)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt  
L.H.  
TITLE An EST database from Sorghum: floral-induced meristems  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 342  
POLYA-No. Location/Qualifiers  
1..375  
/organism="Sorghum propinquum"  
/db\_xref="taxon:132711"

FEATURES  
source  
1..375  
/organism="Sorghum propinquum"  
/db\_xref="taxon:132711"

/clone\_lib="Floral-Induced Meristem 1 (FMI)"  
/note="Organ: Floral-Induced meristems; Vector:  
pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2:  
EcoRI; mature plants were placed in a growth chamber for  
15 days with 16 hr darkness and 8 hr light (flowering is  
induced by short-day conditions); 16 days after being  
returned to the greenhouse under natural long days during  
late April/early May, meristems were harvested The  
library was made from poly-A RNA in the cloning vector  
lambda Zap II. Clones to be sequenced were prepared by  
mass excision."

BASE COUNT 75 a 103 c 133 g 63 t 1 others  
ORIGIN  
Query Match 4.3%; Score 44.8; DB 11; Length 375;  
Best Local Similarity 52.7%; Pred. No. 7.3;  
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY 560 gcagctcttcagctgaagcagctcaggtgtcttctgtcgtgactggggaactcggtga 619  
Db 91 GCAGCAGCGCACCTGGCTGGCGCACAAAGCTGTCTCTGTCTGCGCTCCGACGTCGA 150  
QY 620 ccgcacccaccctgctactcctggttttgaggagatcgctccacagcttctggccaagt 679  
Db 151 CGACCTCGCAAGGTCGACCTACGCGCTGACGTGCTCGCGGCGAGTCAAATCCGATGACAT 210  
QY 680 gttcagctggacagcagcaggtgtcggaggtgttaagtgggtgcgcacatcca 739  
Db 211 GCGGCTCTGTACGAGTGTGCTGGCGGCGGACGCGTACTGGAGATGGACGCGCGCTGCT 270  
QY 740 gccc 743  
Db 271 GGCC 274

RESULT 15  
LOCUS BG560654 379 bp mRNA EST 10-APR-2001  
DEFINITION RH122\_58\_B04.b1\_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA  
sequence.  
ACCESSION BG560654  
VERSION BG560654.1 GI:13589652  
KEYWORDS EST.  
SOURCE Sorghum propinquum.  
ORGANISM Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 379)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt  
L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Seq primer: JEN REV  
High quality sequence stop: 330  
POLYA-No. Location/Qualifiers  
1..379  
/organism="Sorghum propinquum"  
/db\_xref="taxon:132711"

FEATURES  
source  
1..375  
/organism="Sorghum propinquum"  
/clone\_lib="Rhizome2 (RH122)"  
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda  
zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda Zap II.  
Clones to be sequenced were prepared by mass excision."

Search completed: March 7, 2002, 21:07:19  
Job time: 8057 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 21:07:19 ; Search time 1634.39 seconds  
(without alignments)  
5424.206 Million cell updates/sec

Title: US-09-665-728-2\_COPY\_96\_920

Perfect score: 825

Sequence: 1 atgagcctagggcgagct.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_esti:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vit:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	240	29.1	470	10 A1449932	A1449932 mr81b03.x
C 2	90.4	11.0	720	13 A2985073	A2985073 2M0266L17
C 3	77.4	9.4	937	13 CNS025DE	AL181931 Tetraodon
C 4	70.6	8.6	559	11 BF418902	BF418902 UI-R-BJ2-
C 5	64	7.8	915	13 AZ137837	AZ137837 SP_0178_A
C 6	61	7.4	528	10 BE508854	BE508854 del3h07.y
C 7	55	6.7	754	13 AZ189426	AZ189426 SP_1014.A
C 8	46.6	5.6	520	11 BG895162	BG895162 358378 MA
C 9	46.2	5.6	935	13 CNS006XK	AL066051 Drosophil
C 10	46	5.6	965	13 CNS022L5	AL178322 Tetraodon
C 11	45	5.5	818	13 CNS039U2	AL234371 Tetraodon
C 12	44.8	5.4	375	11 BF481079	BF481079 FMI_16_B0

13	44.8	5.4	379	11	BG560654	BG560654 RH1Z2_58
14	44.8	5.4	392	11	BF481089	BF481089 FMI_15_C0
15	44.8	5.4	406	11	BG560698	BG560698 RH1Z2_58
c 16	44.2	5.4	925	13	CNS0091P	AL053013 Drosophil
17	43.2	5.2	523	11	BG240204	BG240204 OVI_19_B0
18	43.2	5.2	924	13	CNS04SRD	AL305554 Tetraodon
19	42.2	5.1	461	10	AU174537	AU174537 AU174537
20	42	5.1	632	11	BG744240	BG744240 602723347
c 21	41.8	5.1	503	10	A1653725	A1653725 WB36F06.x
22	41.8	5.1	686	11	BG835990	BG835990 Zm06_06h0
23	41.8	5.1	925	13	CNS0091P	AL053013 Drosophil
c 24	41.4	5.0	384	10	AA294693	AA294693 SMOV3MCA9
25	41.4	5.0	476	10	BE477180	BE477180 160761 BA
c 26	41.4	5.0	566	10	AW082912	AW082912 xc04c11.x
c 27	41.4	5.0	872	11	BG345072	BG345072 HVSMEG001
28	41	5.0	647	10	BI180123	BI180123 602329707
c 29	41	5.0	778	10	AI991270	AI991270 wu41h01.x
c 30	41	5.0	814	10	AW081400	AW081400 xc41h07.x
c 31	40.8	4.9	523	10	AI855212	AI855212 603010C06
c 32	40.8	4.9	556	11	BG874151	BG874151 MEST46-H0
c 33	40.8	4.9	564	13	AQ848793	AQ848793 LMAJFV1_1
c 34	40.8	4.9	586	10	BE553066	BE553066 946088H08
c 35	40.8	4.9	603	10	AI855218	AI855218 603010D06
c 36	40.8	4.9	626	10	AI855245	AI855245 603010H10
c 37	40.8	4.9	671	13	AQ091789	AQ091789 LMAJFV1_1
38	40.8	4.9	813	11	BG321263	BG321263 Zm04_0390
c 39	40.8	4.9	956	13	CNS03GMI	AL243171 Tetraodon
c 40	40.6	4.9	460	10	AI432376	AI432376 tg54h07.x
c 41	40.6	4.9	738	11	BG864336	BG864336 602798212
42	40.4	4.9	459	10	AW504911	AW504911 UI-HF-BN0
43	40.4	4.9	470	11	BF133055	BF133055 601896455
44	40.4	4.9	472	10	AI330202	AI330202 c4d10ne.f
45	40.4	4.9	503	11	BG754939	BG754939 602711634

ALIGNMENTS

RESULT 1  
A1449932/c

LOCUS mr81b03.xl Stratagene mouse heart (#937316) Mus musculus cDNA clone  
DEFINITION IMAGE:603821 3' similar to WP:F15G9.4 CE01552 IG SUPERFAMILY  
REPEATS:,, mRNA sequence.

ACCESSION A1449932.1 GI:4293462  
VERSION A1449932.1  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 470)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
Possible reversed clone: polyt not found  
High quality sequence stop: 467.  
Location/Qualifiers 1. 470

```
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:603821"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: heart; Vector: pluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
80 a 152 c 168 g 69 t 1 others

BASE COUNT
ORIGIN

Query Match      29.1%; Score 240; DB 10; Length 470;
Best Local Similarity 92.6%; Pred. NO. 1.1e-44;
Matches 252; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 atgacgcctagggcagcactcctgcgctgtctctggcgacctacacagtaagtggcgcg 60
|||||
Db 309 atgacgcctggggcgacgctcctgcgctgtctctggcgacctacacagtaagtggcgcg 250
|||||

Qy 61 gcggtcacatctgtgagccacgaagcgtctgtcccgccacagagagagcaccctg 120
|||||
Db 249 gtgggtcagctgtgagccgcccacgaagacgctgtgtcccgccacagagagagcaccctg 190
|||||

Qy 121 gctctgtcttgatgtcacgggtccatgtgtggagcgtatcgtgagtgatgcagcggc 180
|||||
Db 189 gcttctgtcttgatgtcacgggtccatgtgtggagcgtatcgtgagtgatgcagcggc 130
|||||

Qy 181 gcttcacattctgagcagcagctgtgagcagcagcagcagcagcagcagcagcagc 240
|||||
Db 129 gcttcacattctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 70
|||||

Qy 241 ctggtgcttttcacagccagacatttgccc 272
|||||
Db 69 ctggtgcttttcacagccagacatttgccc 38
|||||

RESULT 2
Az985073/c
LOCUS
DEFINITION
2M0266117R Mouse 10kb plasmid UUC2M library Mus musculus genomic
clone UUC2M0266117 R, DNA sequence.
ACCESSION
Az985073
VERSION
Az985073.1 GI:13856300
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 720)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dduunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
```

```
Plate: 0266 row: L column: 17
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 720.
FEATURES
Location/Qualifiers
1..720
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0266117"
/clone_lib="Mouse 10kb plasmid UUC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gbIAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 203 a 134 c 145 g 238 t
ORIGIN
```

```
Query Match      11.0%; Score 90.4; DB 13; Length 720;
Best Local Similarity 72.0%; Pred. NO. 1.3e-10;
Matches 118; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 331 cagggaggtgactgccagaaatgagtgtgggggccatcaaggtcgctggaggtt 390
|||||
Db 274 CAGGGAGGTGGTGAAGTCCAGAAATGAGCATCGGGGCTATAAAGATTGCCCTGGAAATC 215
|||||

Qy 391 gccaaccccgctcttcacgtctctcgtggtgccagtgcccgctcaccacacaa 450
|||||
Db 214 TCTCTTCCTGGGTCCTTCATCTATGTTTCACTGATGACGATCCAGGATTATCGGCTC 155
|||||

Qy 451 aagaatgagctcctgcagctcctgcagctgaagcagtcgaggt 494
|||||
Db 154 ACCCATGAGGTGCTGCAGCTTATCCACAGAAACAATCTCAGGT 111
|||||

RESULT 3
CNS025DE
LOCUS
DEFINITION
CNS025DE 937 bp DNA GSS 12-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
236N09 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL181931
VERSION
AL181931.1 GI:7820017
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 937)
Roest-Crollius, H., Jaillon, O., Basliva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
REFERENCE
AUTHORS
TITLE
Characterization and repeat analysis of the compact genome of the
```

```

is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: Clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 227-278, >(GAAAA
)n$Simple_repeat
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers
1..559
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bqk-c-06-0-UI"
/clone.lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/note=vector: pT7732-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2

```

143 a 137 c 152 g 127 t

	8.6%;	Score 70.6;	DB 11;	Length 559;
	Similarity 63.0%;	Pred. No. 3.8e-06;		
109; Conservative	0;	Mismatches 64;	Indels 0;	Gaps 0;
agagcgccacctggccttgcgttcctgagtcaaccggctccatctgggacgatctgatg				165
ggggcgctcacgttgctttgtgttgatgtnactggctccatgatgatatttagtt				445
gggtgatcgaggcgccctcacgcatcttgagcgcagctgagocgcgacgcgggtc				225
aggttatTTGAAGGGGCTTCCAAATTTTGAGACGCTCTTTGAAAAGACCTAAGAGACCT				505
cgcgaactatgcgttgtgctttccacgacccagacattggcccagtgac				278

AZ137837 915 bp DNA GSS 28-AUG-2000  
 SP\_0178\_A2\_E10\_SPE6 Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
 genomic clone Plate=178 Col=20 Row=I, DNA sequence.

AZ137837  
 AZ137837.1 GI:8289740  
 GSS.

Strongylocentrotus purpuratus.  
 Strongylocentrotus purpuratus  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euschinoidea; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 915)  
 Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Biondi,T.R.,  
 Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray  
 G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and  
 Hood,L.

A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566

Contact: Cameron, RA, Davidson, EH, Hood, L  
 Division of Biology 156-29  
 California Institute of Technology





ORIGIN	Query Match	5.6%;	Score 46.6;	DB 11;	Length 520;
	Best Local Similarity	51.2%;	Pred. No. 1.1;	Indels	0;
	Matches 109;	Conservative	0;	Mismatches	104;
Qy	13	gcgcagctcctgcgcgtcctgctgacgtacacagtagtggcggcggtcacatct	72		
Db	150	CCCCACCGTGTACGGTGTCTCCGACACCTGCAGCAGCTGTGAGGCTGCCTCTTGAT	209		
Qy	-	73	gatgagcccaagaagcgtgtcccccgcacagagagaccccttgcctcttc	132	
Db	210	GTGCACTTCAACAGGCCAAGTCCCGCTACCAAGAGATGCACACACATAGGCTGC	269		
Qy	133	gatgtcacgcgtccatgtggacgatctgatcgaggtgatcgagcgcgctcacgatt	192		
Db	270	GTCTTCTCGGACCCCATCTTCCAGGAGCTGTGTGGAGCTGAAGCATGCAGCCTCTCGCAC	329		
Qy	193	ctggagcgagctctgagcagcgcgagccgggtc	225		
Db	330	CCATGCACCGTGGCCCGGAGGAGGACGAGGC	362		
RESULT	9				
LOCUS	CNS006XK/c	935 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL066051				
VERSION	AL066051.1	GI:4945019			
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 935)				
TITLE	Genoscope.				
JOURNAL	Direct Submission				
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y; cn bw sp, the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.				
FEATURES	Location/Qualifiers				
source	1. .935				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	/clone_lib="RPCI-98"				
	/clone="BACR14N09"				
	/note="end : T7"				
BASE COUNT	257 a	170 c	162 g	96 t	250 others
ORIGIN					

FEATURES	Location/Qualifiers
source	1. .754
	/organism="Strongylocentrotus purpuratus"
	/db_xref="taxon:7668"
	/clone="plate=1014 Col=11 Row=A"
	/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
	/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"
BASE COUNT	232 a 186 c 124 g 211 t 1 others
ORIGIN	

Query Match	6.7%;	Score 55;	DB 13;	Length 754;
Best Local Similarity	58.1%;	Pred. No. 0.014;		
Matches 97;	Conservative	0;	Mismatches	70;
			Indels	0;
			Gaps	0;
Qy	103	acagagacgcccctcgcttcgtatgacgtccacggtccatgtggagcatttg	162	
Db	363	ACTGGCGAACACGATGGCTTTGTTTGTGATGTAACGGGATCCATGTTGATGACTTG	304	
Qy	163	atgcaggtgatgcgagcgccctacgcattcttgagcgagctgtgagcagccgagccgg	222	
Db	303	CAGCAGGTAATTGAAGGAGCAGAACGGAATTTGGAATTTAATTTGGAGAGGAGAAACAAG	244	
Qy	223	gtcatcgccaactatgctgctggttccttccacgaccagacattgg	269	
Db	243	GCCTTGGAAAAACTTGCTTGGGGGCGCTTCAAGGTTTCAGGGTTGGG	197	
RESULT	8			
LOCUS	BG895162	520 bp	mRNA	EST
DEFINITION	358378 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.			
ACCESSION	BG895162			
VERSION	BG895162.1	GI:14305403		
KEYWORDS	EST.			
SOURCE	pig.			
ORGANISM	Sus scrofa			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
AUTHORS	1 (bases 1 to 520)			
TITLE	Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.			
JOURNAL	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine			
COMMENT	Unpublished (2000) Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@mail.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980504 e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTTCCTCCAGTCACGACG Plate: 121 row: M column: 3 Seq primer: ATTTAGGTGACACTATAG.			
FEATURES	Location/Qualifiers			
source	1. .520			
	/organism="Sus scrofa"			
	/db_xref="taxon:9823"			
	/clone_lib="MARC lPIG"			
	/tissue_type="pooled"			
	/lab_host="DH10B"			
	/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."			
BASE COUNT	95 a	164 c	163 g	98 t

Query Match	5.6%;	Score 46.2;	DB 13;	Length 935;
Best Local Similarity	29.4%;	Pred. No. 1.5;		
Matches 112;	Conservative	90;	Mismatches	178;
			Indels	1;
			Gaps	1;





```

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgtctgactgagggactgcggtga 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 114 GCAGACGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 173

QY 525 ccgcaccacacctgctacactggctttttgagagatcgctccaccagttcttgccaagt 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 CGACCTGCCAAGTCCGACCTACCGCTGACGTGCTCGCGGCAGTCAAAATCCGATGACAT 233

QY 585 gttccagctggaacagcagcaggtgtgtcgagggtgttaaaagtgggtggagtcgcaccatcca 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACCGCGCGTGTCT 293

QY 645 gccc 648
    |||||
Db 294 GGCC 297

RESULT 14
BF481089
LOCUS BF481089 392 bp mRNA EST 05-DEC-2000
DEFINITION FM1_16.C06.bl.A003 Floral-Induced Meristem 1 (FM1) Sorghum
ACCESSION BF481089
VERSION BF481089.1 GI:11551910
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
REFERENCE 1 (bases 1 to 392)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
,L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 350
POLYA=NO.
FEATURES
source Location/Qualifiers
1..392
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FM1)"
/notes="Organ: Floral-Induced meristems; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested The
library was made from poly-A RNA in the cloning vector
lambda ZAP II. Clones to be sequenced were prepared by
mass excision."
BASE COUNT 79 a 108 c 135 g 68 t 2 others
ORIGIN

Query Match 5.4%; Score 44.8; DB 11; Length 392;
Best Local Similarity 52.7%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgtctgactgagggactgcggtga 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 91 GCAGACGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 150

QY 525 ccgcaccacacctgctacactggctttttgagagatcgctccaccagttcttgccaagt 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 CGACCTGCCAAGTCCGACCTACCGCTGACGTGCTCGCGGCAGTCAAAATCCGATGACAT 210

QY 585 gttccagctggaacagcagcaggtgtgtcgagggtgttaaaagtgggtggagtcgcaccatcca 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACCGCGCGTGTCT 270

QY 645 gccc 648
    |||||
Db 271 GGCC 274

RESULT 15
BG560698
LOCUS RHIZ2_58_F04.bl.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
DEFINITION RHIZ2_58_F04.bl.A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
ACCESSION BG560698
VERSION BG560698.1 GI:13589696
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
REFERENCE 1 (bases 1 to 406)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 386
POLYA=NO.
FEATURES
source Location/Qualifiers
1..406
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 90 a 101 c 151 g 64 t
ORIGIN

Query Match 5.4%; Score 44.8; DB 11; Length 406;
Best Local Similarity 52.7%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 465 gcagctctgcagctgaagcagctgcaggggtgtcttctgtctgactgagggactgcggtga 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 GCAGACGCCGACCTGTGCTGGCGCACAAAGCTGTTCTTCTGTCTGCGCTCCGACGTCGA 118

QY 525 ccgcaccacacctgctacactggctttttgagagatcgctccaccagttcttgccaagt 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 CGACCTGCCAAGTCCGACCTACCGCTGACGTGCTCGCGGCAGTCAAAATCCGATGACAT 178

QY 585 gttccagctggaacagcagcaggtgtgtcgagggtgttaaaagtgggtggagtcgcaccatcca 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 GCGCGCTCTGTACGAGTCGCTGGCGGCGGACGCGTACTGGAGATGGACCGCGCGTGTCT 238

QY 645 gccc 648

```

Db 239 GGCC 242  
||||

Search completed: March 7, 2002, 21:07:21  
Job time: 8059 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 7, 2002, 20:48:02 ; Search time 48.12 Seconds  
(without alignments)  
423.320 Million cell updates/sec

Title: US-09-665-728-1  
Perfect score: 1403  
Sequence: 1 MTPRAQLPLLATYTVAA.....IEVRDLGMSQSPPLLMQD 275

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1403	100.0	275	22 AAB83147	Rat secreted facto
2	628.5	44.8	3117	21 AAV53667	Sequence g1/32818
3	116	8.3	353	21 AAV53667	Murine SHD1 protei
4	98	7.0	1017	22 AAB86124	Tobacco formylglyc
5	93	6.6	750	22 AAB48252	Soybean magnesium
6	92.5	6.6	414	13 AAR20202	S.equisimilis stre
7	91.5	6.5	369	12 AAR12892	Truncated Met stre
8	91.5	6.5	372	21 AAV84006	Amino acid sequenc
9	91.5	6.5	384	21 AAV84007	Amino acid sequenc
10	91.5	6.5	401	21 AAV84004	Amino acid sequenc
11	91.5	6.5	413	21 AAV84005	Amino acid sequenc

12	91.5	6.5	414	20 AAB86143	Streptokinase (SK)
13	91.5	6.5	414	20 AAB86144	De-immunised strep
14	91.5	6.5	414	21 AAB01295	Wild type streptok
15	91.5	6.5	414	21 AAB01296	Altered streptokin
16	91.5	6.5	440	12 AAR12889	Streptokinase. St
17	91.5	6.5	483	12 AAR12885	Factor Xa-cleavabl
18	91.5	6.5	483	12 AAR12522	Factor Xa-cleavabl
19	91.5	6.5	499	12 AAR12891	Streptokinase fuse
20	91.5	6.5	747	12 AAR12894	Met-core streptoki
21	91.5	6.5	859	12 AAR12893	OmpAL streptokinas
22	91.5	6.5	1102	19 AAR23948	Porcine phosphoino
23	91.5	6.5	1102	20 AAW90083	Porcine G-protein
24	91.5	6.5	1102	21 AAY76802	Pig p120 regulator
25	90.5	6.5	693	20 AAY30796	A mouse transgluta
26	89.5	6.4	755	22 AAB48248	Corn magnesium che
27	89	6.3	800	18 AAW21723	Plasminogen-bindin
28	89	6.3	813	18 AAW21728	Wild type plasmin
29	89	6.3	1181	18 AAW21727	Streptokinase/malt
30	89	6.3	1194	18 AAW21724	Modified streptoki
31	89	6.3	1194	18 AAW21725	Streptokinase/malt
32	89	6.3	1194	18 AAW21726	Modified streptoki
33	88.5	6.3	404	22 AAG91953	Streptokinase G pr
34	88.5	6.3	440	11 AAR06377	C glutamicum prote
35	88.5	6.3	511	16 AAR69506	Streptokinase G pr
36	88.5	6.3	511	16 AAR69505	Aspergillus sp. re
37	88.5	6.3	515	16 AAR69504	Aspergillus sp. re
38	88.5	6.3	515	16 AAR69504	Aspergillus sp. re
39	88.5	6.3	1407	22 AAB49946	A. oryzae protein
40	88	6.3	848	21 AAB43716	A. thaliana formyl
41	88	6.3	896	17 AAW00733	Human cancer assoc
42	87	6.2	758	19 AAW81771	Human major vault
43	86.5	6.2	348	12 AAR10196	Tobacco CHLD prote
44	86.5	6.2	356	20 AAY24795	Streptokinase (1-3
45	86.5	6.2	372	12 AAR10197	Recombinant strept
					Streptokinase (1-3

ALIGNMENTS

RESULT 1  
AAB83147  
ID AAB83147 standard; protein: 275 AA.  
XX AAB83147;  
XX  
XX 29-JUN-2001 (first entry)  
XX  
XX Rat secreted factor encoded by clone P00210D09.  
XX  
XX Rat; secreted factor: P00210D09; cardiant; nephrotropic;  
KW antiinflammatory; gene therapy; cardiac disease; renal disease;  
KW inflammatory disease.  
XX  
XX Rattus sp.  
XX  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..21  
XX Protein /label= Signal\_peptide  
XX /note= "Rat secreted factor"  
XX Domain 35..55  
XX /label= Transmembrane\_domain  
XX Domain 123..143  
XX /label= Transmembrane\_domain  
WO200123419-A2.  
05-APR-2001.  
27-SEP-2000; 2000WO-US26582.  
27-SEP-1999; 99US-0156277.

PA (SCIO-) SCIOS INC.  
 XX Stanton LW, Kapoun AM;  
 XX WPI; 2001-328177/34.  
 XX N-PSDB; AAF82464.  
 XX Novel secreted factor encoded by clone P00210D09 useful for diagnosing,  
 PT treating and/or preventing various cardiac, renal and inflammatory  
 PT diseases -  
 XX  
 XX Claim 9; Fig 1; 69pp; English.  
 XX  
 CC The present sequence is a novel secreted factor encoded by rat cDNA  
 CC clone P00210D09. The invention relates to a polypeptide comprising a  
 CC sequence of at least 80% identity to residues 22-122 of the present  
 CC sequence, or a sequence encoded by a nucleic acid hybridizing under  
 CC stringent conditions to the complement of the coding region comprising  
 CC 1031 nucleotides, and having at least one biological activity of the  
 CC polypeptide encoded by clone P00210D09. The polypeptides and  
 CC polynucleotides of the invention are useful for the treatment of  
 CC cardiac, renal and inflammatory diseases. The polynucleotides are  
 CC useful in antisense mediated gene inhibition and in gene therapy.  
 CC The polypeptides are useful in assays for identifying lead compounds  
 CC that may be used as therapeutic agents in the treatment of cardiac,  
 CC kidney or inflammatory diseases.  
 XX  
 SQ Sequence 275 AA;  
 Query Match 100.0%; Score 1403; DB 22; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-138;  
 Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTPRAQLLLLATYVWAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVIDG 60  
 Db 1 mtpraqlllllatyvcvaavtsdeptkltspatgdtalafvfdvtgsmwddlmqvldg 60  
 QY 61 ASRILERSLSRSRVIANVALYVDFHDPDTPVLTADPVVFORLRLQLYVGGGDCPEMS 120  
 Db 61 asrilerslsrsrvianvalyvfhdpgvltadpvpvfqrelrqlvvgggdcpems 120  
 QY 121 VCAIKAAVEANPGSFIYVFSARAKDYHKKNELQLQLKQSQVVFVLTGCGDRTHPG 180  
 Db 121 vgaikaaveanpgsfliyfsdarakdyhknellqlqlkqsgvvfvltgdcgdrthpg 180  
 QY 181 YIAFEBIATSSGOVFQDKQVSEVLKQVSAIOASKVHLLSADHEEGEHTWRIPFDP 240  
 Db 181 ylafebiaatssgqvfdkqkvsevlkwvesaioaskvhllsadheeegehtwrifpdp 240  
 QY 241 SLKEVTISLGGPGEIEVRDPLGMSQGSPLLMQD 275  
 Db 241 slkevtislsgrpgeievrplgmsqgspllmqd 275  
 RESULT 2  
 AAY53667  
 ID AAY53667 standard; Protein; 3117 AA.  
 XX  
 AC AAY53667;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE Sequence gi/3328186 from an alignment with protein 608.  
 XX  
 KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;  
 KW bone development; gi/3328186.  
 XX  
 OS Unidentified.  
 OS  
 PN WO9960164-A1.  
 XX  
 DT 25-NOV-1999.  
 PD

XX 14-MAY-1999; 99WO-US11066.  
 XX 15-MAY-1998; 98US-0085673.  
 XX (QUAR-) QUARK BIOTECH INC.  
 XX Einat P, Mor O, Skalliter R, Feinstein E, Faerman A;  
 XX WPI; 2000-053304/04.  
 XX Identification of stress induced genes for determining risk and  
 PT preventing, treating or controlling osteoporosis -  
 PT  
 XX Claim 32; Fig 6A-R; 308pp; English.  
 XX  
 CC The present sequence is obtained from a clustral X alignment with  
 CC protein 608. Protein 608 was identified using the method of the invention  
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the  
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected  
 CC to mechanical strain. The specification describes a method for the  
 CC identification of genes responsive to a specific mechanical stress. The  
 CC method comprises applying the mechanical stress to an organism (tissue  
 CC or cells comprising bone cells), isolating the specific cellular  
 CC fractions and extracting mRNA from them, and differentially analysing the  
 CC mRNA in comparison with control samples. The method is used to identify  
 CC genes whose expression is responsive to a specific stress. The identified  
 CC genes are employed in determining risk associated with a physiological or  
 CC disease state. The risk determination methods are used for testing a  
 CC medicament for gene therapy. These medicaments, or genes identified by  
 CC the method of the invention, are used for treating, preventing or  
 CC controlling a physiological or disease state (especially osteoporosis or  
 CC bone density or other factors causing or contributing to osteoporosis or  
 CC its symptoms or other conditions involved in mechanical stress or its  
 CC lack. The methods can also be used for advancing research or studies in  
 CC bone development.  
 XX  
 SQ Sequence 3117 AA;  
 Query Match 44.8%; Score 628.5; DB 21; Length 3117;  
 Best Local Similarity 48.6%; Pred. No. 9.5e-56;  
 Matches 126; Conservative 51; Mismatches 69; Indels 13; Gaps 3;  
 QY 10 LLATYVVAAVTSDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVTDGASRILERSL 69  
 Db 15 lllatvvaavtsdeptkltspatgdtalafvfdvtgsmwddlmqvtdgasklftvm 66  
 QY 70 SRSRVIANKALVPEHDPDTPVLTADPVVFORLRLQLYVGGGDCPEMSVGAATKAAVE 129  
 Db 67 adrekliynyimvpfhdpylgelintctdtyfmrqlskvyvhggdcpektitgllkaiq 126  
 QY 130 VANPGSFIYVFSARAKDYHKKNELQLQLKQSQVVFVLTGCGDRTHPGYLAPEEIAS 189  
 Db 127 islpssfiyftdarskdylhedevalntiqekssvfvmtgdcgnrthpgftrtyekiaa 186  
 QY 190 TSSGOVFQDKQVSEVLKQVSAIOASKVHLLSADHEEGEHTWRIPDPSELKEVTISL 249  
 Db 187 asfggvfhlekadvstvleyvrhvkqkvhlmyearerggtvsvrnlpvdkhnlseitia 246  
 QY 250 SG-----PGPEIEVRDPLG 263  
 Db 247 sgkdksdndldiviropeg 265  
 RESULT 3  
 AAB98973  
 ID AAB98973 standard; protein; 353 AA.  
 XX  
 AC AAB98973;  
 XX  
 DT 20-AUG-2001 (first entry)  
 XX



DE Murine SHD1 protein.

KW Mouse; SHD1; differentiation; germ centre B cell.

XX Mus sp.

XX JP2001078778-A.

XX 27-MAR-2001.

XX 17-SEP-1999; 99JP-0263688.

XX 17-SEP-1999; 99JP-0263688.

XX (SUME ) SUMITOMO ELECTRIC IND CO.

XX WPI; 2001-313371/33.

DR N-PSDB; AAH25841.

XX SHD1 protein useful for controlling the differentiation and activation of germ center B cells

XX Claim 1; Page 6-7; 16pp; Japanese.

XX The present invention relates to the murine SHD1 protein shown here. This protein can be used for controlling differentiation and activation of germ centre B cells.

XX Sequence 353 AA;

Query Match 8.3%; Score 116; DB 22; Length 353;  
Best Local Similarity 24.9%; Pred. No. 0.0015;  
Matches 70; Conservative 31; Mismatches 98; Indels 82; Gaps 13;

QY 3 PRAQLL---PLLATYTVVAAVTSDEPTKTLSPATGDATLA----FVFDVTGSMMDL- 54  
DB 64 ppsllrppvllatvrylageva-----grgdvscaevsfvadravrldls 113  
QY 55 MQVID--CASRTILERSLSRSRVIANALVPFDPDGPVIL--TADPWVQRE----- 104  
DB 114 lqgvdaadaatvleaalllavvar-----vrpeetrgaadpvlitqvgefg 163  
QY 105 -LRQLYVOGGGDCP-----EMSVGAIKAAVEVANPGSFIVVFSARAK 146  
DB 164 slrrcyargkgyprqaafqglfllynlgslapaggstaalracpqlqaalavdaafre 223  
QY 147 DYHKK-NELLQLLQSQWVFLTGDCGDRTHPGYLAFEEIA-----STSSGQVFOLD 199  
DB 224 dnharlfrllrtlpylqs-----cavqehigyarrkalarlsralstpkgtlpld 274  
QY 200 KOQVSEVLKWEVSAIQASKVHLLSAD-----HEEEG 230  
DB 275 fiehflaidglqearldcqhgltdkdrrvflrgysee 315

RESULT 4

AA886124

ID AAB86124 standard; Protein; 1017 AA.

XX AAB86124;

XX 26-JUL-2001 (first entry)

XX Tobacco formylglycinamide-ribotide synthase protein fragment.

XX Plant; formylglycinamide-ribotide synthase; E.C. 6.3.5.3; herbicide;  
KW purine biosynthesis; defoliant; desiccant; transgenic.

XX Nicotiana tabacum.

OS WO200131025-A2.

XX

PD 03-MAY-2001.

XX 17-OCT-2000; 2000WO-EP10204.

XX 25-OCT-1999; 99DE-1051443.

XX (BADI ) BASF AG.

XX Lerchl J, Ehrhardt T, Sonnewald U, Boldt R, Kunze G;

XX WPI; 2001-316337/33.

DR N-PSDB; AAF88175.

XX New DNA encoding formylglycinamide-ribotide synthase of plants, useful for identifying enzyme inhibitors, potential herbicides -

XX Example 2; Page 54-58; 60pp; German.

XX This invention describes novel polynucleotide sequences from Arabidopsis thaliana, tobacco (Nicotiana tabacum) and Chilopsis linearis which contain the coding region for a plant formylglycinamide-ribotide synthase (E.C. 6.3.5.3), an enzyme used in purine biosynthesis. The products of the invention have herbicide, defoliant and desiccant activity. Plant formylglycinamide-ribotide synthase can be expressed in prokaryotic or eukaryotic cells, e.g. to produce transgenic plants with increased resistance to herbicides that target formylglycinamide-ribotide synthase and to prepare a test system for identifying inhibitors of formylglycinamide-ribotide synthase. Such inhibitors are potentially useful as defoliants, desiccants or especially herbicides, either total or selective, depending on application rate. This sequence represents the tobacco formylglycinamide-ribotide synthase described in the method of the invention.

XX Sequence 1017 AA;

Query Match 7.0%; Score 98; DB 22; Length 1017;  
Best Local Similarity 19.9%; Pred. No. 0.56;  
Matches 57; Conservative 52; Mismatches 94; Indels 84; Gaps 12;  
QY 10 LLLATYTVVAAVTSDEPTKTLSP-----ATGDTATLAFVDFVTGS 49  
DB 512 lvisy-----vtcpditktvtplkigddgvlhldlakgrrlggsalagvdfdigd 565  
QY 50 MWDDLQVIDGASRTILERSLSRSRVIANALVPFDPDGPVTLTADPVPVFORELRLQY 109  
DB 566 espdladv-----syktvfvneqvnlisdelisaghdisdggllvnaemaf----- 612  
QY 110 VOGGDCPEMSVGAIKAAVEVANPGSFI--YVFSARAKDYHKRNLQLQLKQSQVVF 167  
DB 613 ---agnc-----gihlditslgssvptlfaf-----elgllievsknldl 651  
QY 168 VLTGDCGDRTHPGYLAFEEIASTSSGQVFOQL-----DKQVSEVLKWEVSAIQASK 218  
DB 652 vlekics-----gavsanliqvtsspivelrvdgvthlnktsvlrdm-weetsfqlek 705  
QY 219 VHLLSA--DHEEEG-----EHTWRIPFDPSPSLKEVTSLSGPGPEIEV 258  
DB 706 lqlascvelekeglknrhepswklsftptftddkymtavskpkvai 752

RESULT 5

AA848252

ID AAB48252 standard; Protein; 750 AA.

XX AAB48252;

XX 02-APR-2001 (first entry)

XX Soybean magnesium chelataase subunit (clone sdp4c.pk022.h18[FIS]).  
DE Magnesium chelataase; transgenic; herbicide; gene marker; plant breeding;  
KW soybean.

XX OS Glycine max.  
XX PN WO200075340-A2.  
XX PD 14-DEC-2000.  
XX PF 02-JUN-2000; 2000WO-US15351.  
XX PR 04-JUN-1999; 99US-0137461.  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX PI Butler KH, Famodu OO, Gutteridge S, Maxwell CA;  
XX DR N-PSDB; AAC84571.  
XX PT Isolated nucleic acid fragments encoding magnesium chelatase subunits,  
XX PF useful as probes for genetic and physical mapping of genes, as markers  
XX PT for traits linked to these genes, and in plant breeding -  
XX PS Claim 10; Page 60-62; 103pp; English.  
XX CC The invention relates to nucleic acid fragments encoding magnesium  
XX CC chelatase subunits. The nucleic acid fragments may be used to create  
XX CC transgenic plants in which the new polypeptides are present at higher or  
XX CC lower levels than normal or in cell types or developmental stages in  
XX CC which they are not normally found, and for overexpression in bacterial  
XX CC or yeast hosts to efficiently produce large amounts of the encoded  
XX CC polypeptides which could then be used for screening different compounds  
XX CC for potential herbicidal activity. The polynucleotides may also be used  
XX CC as probes for genetic and physical mapping the genes that they are part  
XX CC of, and as markers for traits linked to these genes. Such information  
XX CC is useful in plant breeding. The polypeptides are used for preparing  
XX CC antibodies, which are useful for detecting the polypeptides in situ or  
XX CC in vitro, and as a target to facilitate design and/or identification of  
XX CC inhibitors of enzymes that may be used as herbicides. Host cells may also  
XX CC be used directly for screening different compounds for potential  
XX CC herbicidal activity. The present sequence represents the soybean  
XX CC magnesium chelatase subunit.  
XX SQ Sequence 750 AA;

Query Match 6.6%; Score 93; DB 22; Length 750;  
Best Local Similarity 23.0%; Pred. No. 1.2;  
Matches 54; Conservative 38; Mismatches 85; Indels 58; Gaps 11;  
QY 24 SDEPTKTLSPATGDTAFVFDVTGSMWDDLMQVIDGAS-RILERSLSRSRVIANALV 82  
DB 533 tdmrakrmakag-alvifvdasgsmalnrmqnakgaalkilaesytardqv---sll 587  
QY 83 PFHDPDGPVTL--TADPVVFORQLYVQGGDCCPEMSVGAIAAEEVANPFS---F 136  
DB 588 pfirg-daaevllppsrslamarkrierlpcggsgplahgttavrvglnaeksgdvgvrm 646  
QY 137 IYVFSARAK-----DYHK-----KNELLQL-----LQLKQSQV 165  
DB 647 lvaiddgranislrstbdveaatdapksaqelkdellelvagkiyagmsllvldten 706  
QY 166 VFVLGDCGDRTHPGYLAPEEIAS--TSSGQVFDLQKQVSEVLKWEESAIQAS 217  
DB 707 kfvtstgfa-----kelarvaqgkyylpnasdavissatkealsalkss 750

RESULT 6  
AAR20202  
ID AAR20202 standard; Protein; 414 AA.  
AC AAR20202;  
XX AAR20202;  
DT 21-APR-1992 (first entry)

XX DE S.equisimilis streptokinase.  
XX KW plasminogen activator; coronary thrombosis; ATCC-9542.  
XX OS Streptococcus equisimilis.  
XX PN AU9178101-A.  
XX PD 28-NOV-1991.  
XX PF 31-MAY-1991; 91AU-0078101.  
XX PR 23-MAY-1990; 90CU-00000090.  
XX PA (INGI-) CENT ING GENETICA B.  
XX PI Garcia MPE, Chaplen RR, Hidalgo AP, Doce RS, Marrero LFH;  
XX PI Collazo PR, Ramirez AC, Munoz EAM, Martinez WB, Somavilla MC;  
XX PI Fernandez AP, Garcia J, Martinez LSH;  
XX WPI: 1992-024716/04.  
XX DR N-PSDB; AAQ20665.  
XX PT Streptokinase C-2 gene from S.equisimilis type C - plasmids and  
XX PT transformants for large scale intra- and extracellular expression  
XX PT of SKC-2 useful in thrombolytic agents  
XX PS Claim 13; Page 14; 28pp; English.  
XX CC The SKC-2 gene was isolated from S.equisimilis type C by gene  
XX CC amplification from the synthetic primers SK1, SK2 and SK3 (see  
XX CC AAQ20666-8). The SKC-2 gene was inserted into yeast expression vector  
XX CC pPS-7 for extracellular expression of streptokinase and into pNAO  
XX CC for intracellular expression in yeast. For expression in bacteria,  
XX CC the SKC-2 gene was inserted in plasmid pBK3, between a trp promoter  
XX CC and a T4 terminator.  
XX SQ Sequence 414 AA;

Query Match 6.6%; Score 92.5; DB 13; Length 414;  
Best Local Similarity 25.5%; Pred. No. 0.54;  
Matches 56; Conservative 33; Mismatches 90; Indels 41; Gaps 10;  
QY 52 DDLMOVIDGASRILERSLSRSRVIANALVPFHDPIGPTLTADPVVFORQLYVQ 111  
DB 95 ddyfveidfas---datidrn-----gkvyfadkd-gsvtlptqpv-----geffls 138  
QY 112 GG---GDCPEMSVGAIAAEEVANPFSFIYFSDAKADYHKKNELLQLQLKQ---SQV 165  
DB 139 ghvrvrpykekpigpqaksdvveytvqfcplopddfrpgldkdkllktlaigdttsge 198  
QY 166 VFVLGDCGDRTHPGYLAPEEIAS--TSSGQVFDLQKQVSEVLKWEESAIQASVHLLS 223  
DB 199 llaqaqslinlkthpdytyierdsivthndifrtilpmdqgeftyhvkneqayeinkks 258  
QY 224 ADHEE-----EGEHTWRIPDPS-LKEVTI 247  
DB 259 gineeinntdlisekyyvllkgekpyd-pfdrshklkfti 297

RESULT 7  
AAR12892  
ID AAR12892 standard; Protein; 369 AA.  
AC AAR12892;  
XX AAR12892;  
DT 17-SEP-1991 (first entry)  
XX Truncated Met streptokinase.  
XX DE Fusion protein; blood clotting; coagulation; fibrinolysis;  
KW

KW antithrombotic; thrombolysis; streptokinase; thrombin.  
 XX  
 OS Streptococcus equisimilis ATCC 9542 or ATCC 10009.  
 XX

Key Location/Qualifiers  
 FT Protein 2..369  
 FT /label= core streptokinase  
 FT /note= "AAs 16-383"  
 XX

PN W09109125-A.

XX 27-JUN-1991.

XX 07-DEC-1990; 90WO-GB01911.

XX 07-DEC-1990; 90WO-GB01911.

PR 07-DEC-1989; 89GB-0027722.

XX (BRBI-) BRIT BIO-TECHN LTD.

XX Dawson KM, Hunter MG, Czaplowski LG;

DR WPI; 1991-208151/28.

DR N-PSDB; AAQ12159.

XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 XX fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis.  
 PT  
 XX

PS Disclosure; Page 88; 115pp; English.

XX The sequence was deduced from DNA obtd. from PCR amplified chromo-  
 CC somal DNA from S. equisimilis (Lancefield's Gp C.) ATCC 10009 or  
 CC ATCC 9642. The gene was truncated by 15 codons at the 3' ter-  
 CC minial and engineered to add a Met codon prior to codon 16 of the  
 CC streptokinase CDS. The construct was used to prepare expression  
 CC vector pGC606 for prodn. of the streptokinase core molecule.  
 CC See also AAR12887-R12889, AAR12891-R12893, AAR12885 and AAR12522.  
 XX

XX Sequence 369 AA;

Query Match 6.5%; Score 91.5; DB 12; Length 369;  
 Best Local Similarity 25.3%; Pred. No. 0.57;  
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLMOVIDGASRIILERSLSRSRVIANVALVFPFHPDIPGVTLTADPV-----VFQR 103

Db 81 ddyfevidfas---datitdrn-----gkvyfadkd-gsvtlptqpvefllsgghrvr 130

QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEANPGSFYVFSARAKDYHKKNELLQLKQ- 162

Db 131 pykekpqnqa-----ksvdvetyvtqfplpdpddfrpglkdtkllkclgdt 179

QY 163 --SQVFLVLTGDCDTRHPGYLAPEEIAS--TSSQVQFQDKQVSEVSKWVESAIQASK 218

Db 180 itsqellaqagsilnkthpgygtiyerdssivthdndifrtilpmdqeftyhvkneqaye 239

QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247

Db 240 inkksglneeinntdlisekyvylkqekpyd-pfdrshlklfti 283

RESULT 8

AY84006

XX AAY84006 standard; Protein; 372 AA.

XX AAY84006;

XX 03-JUL-2000 (first entry)

XX Amino acid sequence of a mutant streptokinase.

DE

XX

KW Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;  
 KW plasminogen activator complex formation; thrombolytic;  
 KW myocardial infarction; pulmonary thromboembolism; thrombosis.  
 XX  
 OS Synthetic.  
 OS Streptococcus equisimilis.  
 XX  
 PN EP985729-A2.  
 XX  
 XX 15-MAR-2000.

XX 13-AUG-1999; 99EP-0202639.

XX 14-AUG-1998; 98CU-0000119.

XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.

XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;

PI Seralena Menendez A;

DR WPI; 2000-226041/20.

DR N-PSDB; AA299251.

XX Recombinant production of streptokinase mutants, useful as thrombolytic  
 PT agents for treating myocardial infarction, that are truncated at the N-  
 PT or C-terminus  
 XX  
 XX Disclosure; Page 22-23; 54pp; English.

XX The present sequence represents a mutant protein of the streptokinase  
 CC SKC-2. Streptokinase forms a complex with plasminogen, activating its  
 CC conversion to plasmin. The SKC-2 gene was modified to produce the  
 CC mutant protein. The obtained mutants conserve their capacity for  
 CC plasminogen activator complex formation, thus having reduced  
 CC antigenicity. The mutants are derived from the 1-119 gene fragment,  
 CC and retain the thrombolytic activity of SKC-2 but, compared with the  
 CC full-length protein, are less antigenic and retain activity better in  
 CC presence of antibodies that neutralize activity of the complete  
 CC protein. The mutants are thrombolytic agents which are useful for  
 CC treating myocardial infarction, pulmonary thromboembolism, surgical  
 CC complications and other forms of thrombosis.  
 XX

XX Sequence 372 AA;

Query Match 6.5%; Score 91.5; DB 21; Length 372;  
 Best Local Similarity 25.3%; Pred. No. 0.58; Indels 51; Gaps 10;  
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;

QY 52 DDLMOVIDGASRIILERSLSRSRVIANVALVFPFHPDIPGVTLTADPV-----VFQR 103

Db 95 ddyfevidfas---datitdrn-----gkvyfadkd-gsvtlptqpvefllsgghrvr 144

QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEANPGSFYVFSARAKDYHKKNELLQLKQ- 162

Db 145 pykekpqnqa-----ksvdvetyvtqfplpdpddfrpglkdtkllkclgdt 193

QY 163 --SQVFLVLTGDCDTRHPGYLAPEEIAS--TSSQVQFQDKQVSEVSKWVESAIQASK 218

Db 194 itsqellaqagsilnkthpgygtiyerdssivthdndifrtilpmdqeftyhvkneqaye 253

QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247

Db 254 inkksglneeinntdlisekyvylkqekpyd-pfdrshlklfti 297

RESULT 9

AY84007

XX AAY84007 standard; Protein; 384 AA.

XX AAY84007;

XX 03-JUL-2000 (first entry)

XX

XX Amino acid sequence of a mutant streptokinase.  
DE Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;  
XX plasminogen activator complex formation; thrombolytic;  
KW myocardial infarction; pulmonary thromboembolism; thrombosis.  
XX Synthetic.  
OS Streptococcus equisimilis.  
XX EP985729-A2.  
PN 15-MAR-2000.  
XX 13-AUG-1999; 99EP-0202639.  
PF 14-AUG-1998; 98CU-0000119.  
XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.  
XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;  
PI Seralena Menendez A;  
XX WPI; 2000-226041/20.  
DR N-PSDB; AAZ99252.  
XX Recombinant production of streptokinase mutants, useful as thrombolytic  
PT agents for treating myocardial infarction, that are truncated at the N-  
PT of C-terminus -  
XX Disclosure; Page 24-25; 54pp; English.  
PS The present sequence represents a mutant protein of the streptokinase  
XX SKC-2. Streptokinase forms a complex with plasminogen, activating its  
CC conversion to plasmin. The SKC-2 gene was modified to produce the  
CC mutant protein. The obtained mutants conserve their capacity for  
CC plasminogen activator complex formation, thus having reduced  
CC antigenicity. The mutants are derived from the 1-1119 gene fragment,  
CC and retain the thrombolytic activity of SKC-2 but, compared with the  
CC full-length protein, are less antigenic and retain activity better in  
CC presence of antibodies that neutralize activity of the complete  
CC protein. The mutants are thrombolytic agents which are useful for  
CC treating myocardial infarction, pulmonary thromboembolism, surgical  
XX complications and other forms of thrombosis.  
XX Sequence 384 AA;  
SQ  
Query Match 6.5%; Score 91.5; DB 21; Length 384;  
Best Local Similarity 25.3%; Pred. No. 0.61;  
Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;  
QY 52 DDLQVIGASRLERSLSRSRVIANVALVPHDPDTPVTLTADPV-----VFQR 103  
DB 95 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqgefllghvr 144  
QY 104 ELRLQYVGGGDCPEMSVGAITAAVEVANPGSFYVFSDDAKDYHKKNELLLQLKQ- 162  
DB 145 pykekpignqa-----ksvdyevyvtqftplnpdddfpgrlkdtklkltaigt 193  
QY 163 --SQVFFVLTGCDGDRTHPGYLAPEEIAS--TSSGOVFDLQKQVSEVLKWWESAIOASK 218  
DB 194 itsqellaqaqsilnkthpgytiyerdssivthdndifrlpmdqefthyvknreqaye 253  
QY 219 VHLLSADHEE-----EGEHTWRIFPDPS-LKEVTI 247  
DB 254 inkksglneenintdlisekyyvikkgkpyd-pfdrshlklfti 297  
RESULT 10  
RAY84004  
ID AAY84004 standard; Protein; 401 AA.  
XX

AC AAY84004;  
XX 03-JUL-2000 (first entry)  
DE Amino acid sequence of a mutant streptokinase.  
XX Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;  
KW plasminogen activator complex formation; thrombolytic;  
KW myocardial infarction; pulmonary thromboembolism; thrombosis.  
XX Synthetic.  
OS Streptococcus equisimilis.  
XX EP985729-A2.  
PN 15-MAR-2000.  
XX 13-AUG-1999; 99EP-0202639.  
PF 14-AUG-1998; 98CU-0000119.  
XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.  
XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;  
PI Seralena Menendez A;  
XX WPI; 2000-226041/20.  
DR Recombinant production of streptokinase mutants, useful as thrombolytic  
PT agents for treating myocardial infarction, that are truncated at the N-  
PT of C-terminus -  
XX Disclosure; Page 18-20; 54pp; English.  
PS The present sequence represents a mutant protein of the streptokinase  
XX SKC-2. Streptokinase forms a complex with plasminogen, activating its  
CC conversion to plasmin. The SKC-2 gene was modified to produce the  
CC mutant protein. The obtained mutants conserve their capacity for  
CC plasminogen activator complex formation, thus having reduced  
CC antigenicity. The mutants are derived from the 1-1119 gene fragment,  
CC and retain the thrombolytic activity of SKC-2 but, compared with the  
CC full-length protein, are less antigenic and retain activity better in  
CC presence of antibodies that neutralize activity of the complete  
CC protein. The mutants are thrombolytic agents which are useful for  
CC treating myocardial infarction, pulmonary thromboembolism, surgical  
XX complications and other forms of thrombosis.  
XX Sequence 401 AA;  
SQ  
Query Match 6.5%; Score 91.5; DB 21; Length 401;  
Best Local Similarity 25.3%; Pred. No. 0.65;  
Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;  
QY 52 DDLQVIGASRLERSLSRSRVIANVALVPHDPDTPVTLTADPV-----VFQR 103  
DB 82 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqgefllghvr 131  
QY 104 ELRLQYVGGGDCPEMSVGAITAAVEVANPGSFYVFSDDAKDYHKKNELLLQLKQ- 162  
DB 132 pykekpignqa-----ksvdyevyvtqftplnpdddfpgrlkdtklkltaigt 180  
QY 163 --SQVFFVLTGCDGDRTHPGYLAPEEIAS--TSSGOVFDLQKQVSEVLKWWESAIOASK 218  
DB 181 itsqellaqaqsilnkthpgytiyerdssivthdndifrlpmdqefthyvknreqaye 240  
QY 219 VHLLSADHEE-----EGEHTWRIFPDPS-LKEVTI 247  
DB 241 inkksglneenintdlisekyyvikkgkpyd-pfdrshlklfti 284  
RESULT 11  
RAY84005

ID XX AAY84005 standard; Protein; 413 AA.  
 AC AAY84005;  
 XX 03-JUL-2000 (first entry)  
 DT XX  
 DE XX Amino acid sequence of a mutant streptokinase.  
 DE XX  
 DE XX Streptokinase; SKC-2; plasminogen; plasmin; antigenicity;  
 KW plasminogen activator complex formation; thrombolytic;  
 KW myocardial infarction; pulmonary thromboembolism; thrombosis.  
 XX XX  
 OS Synthetic.  
 OS Streptococcus equisimilis.  
 XX EP985729-A2.  
 XX 15-MAR-2000.  
 XX 13-AUG-1999; 99EP-0202639.  
 XX 14-AUG-1998; 98CU-0000119.  
 XX (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.  
 XX Torrens Madrazo IDC, Garcia Ojalvo A, De La Fuente Garcia JDJ;  
 PI Seralena Menendez A;  
 XX WPI; 2000-226041/20.  
 DR N-PSDB; AAZ99250.  
 XX Recombinant production of streptokinase mutants, useful as thrombolytic  
 PT agents for treating myocardial infarction, that are truncated at the N-  
 PT or C-terminus  
 XX Disclosure; Page 20-21; 54pp; English.  
 XX The present sequence represents a mutant protein of the streptokinase  
 CC SKC-2. Streptokinase forms a complex with plasminogen, activating its  
 CC conversion to plasmin. The SKC-2 gene was modified to produce the  
 CC mutant protein. The obtained mutants conserve their capacity for  
 CC plasminogen activator complex formation, thus having reduced  
 CC antigenicity. The mutants are derived from the 1-1119 gene fragment,  
 CC and retain the thrombolytic activity of SKC-2 but, compared with the  
 CC full-length protein, are less antigenic and retain activity better in  
 CC presence of antibodies that neutralize activity of the complete  
 CC protein. The mutants are thrombolytic agents which are useful for  
 CC treating myocardial infarction, pulmonary thromboembolism, surgical  
 CC complications and other forms of thrombosis.  
 XX Sequence 413 AA;  
 SQ  
 Query Match 6.5%; Score 91.5; DB 21; Length 413;  
 Best Local Similarity 25.3%; Pred. No. 0.68;  
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;  
 QY 52 DDLMOVIGASRILERSLSRSRVIANVALYFPHDPDGPVTLTADPV-----VFQR 103  
 DB 82 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqefllsgghrvr 131  
 QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEVANPGSFYVFSARAKDYHKKNELLLQLLQKQ- 162  
 DB 132 pykekpqnqa-----ksvdyeyvtqftplnpddfrpglkdtklktlaigt 180  
 QY 163 --SQVFLVIGDGRTHPGYLAPEFIAS--TSSQVQLDKQVSEVLKWWVESAIQASK 218  
 DB 181 itsqellaqaqsilnktbpgtyiyerdssivthdndifrtlpmdqeftyhvkneqaye 240  
 QY 219 VHLISADHEE-----EGEHTWRIPDPDS-LKEVTI 247  
 DB 241 inkksglneecintdlisekyyvkkgekyd-pfrdshlklftl 284

RESULT 12  
 AAW86143  
 ID AAW86143 standard; Protein; 414 AA.  
 XX  
 AC AAW86143;  
 XX 03-MAR-1999 (first entry)  
 DT XX  
 DE XX Streptokinase (SK) protein sequence.  
 DE XX  
 KW Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;  
 KW Immunogl bulin; therapeutic; streptokinase.  
 XX  
 OS Streptococcus equisimilis.  
 XX WO9852976-A1.  
 XX 26-NOV-1998.  
 XX 21-MAY-1998; 98WO-GB01473.  
 XX 14-APR-1998; 98GB-0007751.  
 PR 21-MAY-1997; 97GB-0010480.  
 PR 31-JUL-1997; 97GB-0016197.  
 PR 28-NOV-1997; 97GB-0025270.  
 PR 02-DEC-1997; 97US-0067235.  
 XX (BIOV-) BIOVATION LTD.  
 PA Carr FJ;  
 XX WPI; 1999-045301/04.  
 DR  
 XX Reducing immunogenicity of proteins - by modifying the amino acid  
 PT sequence of the protein to eliminate potential epitopes for T-cells  
 PT of a given species  
 XX Example 6; Fig 28; 77pp; English.  
 XX The invention relates to a method for the production of non-immunogenic  
 CC proteins. The method comprises determining at least part of the amino  
 CC acid sequence of the protein; (b) identifying in the amino acid sequence  
 CC one or more potential epitopes for T-cells (T-cell epitopes) of the given  
 CC species; and (c) modifying the amino acid sequence to eliminate at least  
 CC one of the T-cell epitopes identified in step (b) thereby to eliminate or  
 CC reduce the immunogenicity of the protein when exposed to the immune  
 CC system of the given species. A method of analysing a pre-existing protein  
 CC to predict the basis for immunogenic responses is also provided. The  
 CC methods can be used particularly for reducing the immunogenicity of  
 CC immunoglobulins or therapeutic proteins, e.g. streptokinase (SK). The  
 CC products can be used for diagnosis and therapy. The present sequence  
 CC represents the amino acid sequence of the SK protein.  
 XX Sequence 414 AA;  
 SQ  
 Query Match 6.5%; Score 91.5; DB 20; Length 414;  
 Best Local Similarity 25.3%; Pred. No. 0.68;  
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;  
 QY 52 DDLMOVIGASRILERSLSRSRVIANVALYFPHDPDGPVTLTADPV-----VFQR 103  
 DB 95 ddyfevidfas---datitdrn-----gkyvfadkd-gsvtlptqpvqefllsgghrvr 144  
 QY 104 ELRQLYVGGGDCPEMSVGAIKAAVEVANPGSFYVFSARAKDYHKKNELLLQLLQKQ- 162  
 DB 145 pykekpqnqa-----ksvdyeyvtqftplnpddfrpglkdtklktlaigt 193  
 QY 163 --SQVFLVIGDGRTHPGYLAPEFIAS--TSSQVQLDKQVSEVLKWWVESAIQASK 218  
 DB 194 itsqellaqaqsilnktbpgtyiyerdssivthdndifrtlpmdqeftyhvkneqaye 253

QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247  
 Db 254 inksglneeinntdlisekyvllkgekpyd-pfdrshklfti 297

RESULT 13

AAW86144  
 ID AAW86144 standard; Protein; 414 AA.  
 AC AAW86144;  
 XX  
 XX 03-MAR-1999 (first entry)  
 XX De-Immunised streptokinase (SK) protein sequence.  
 XX Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;  
 KW immunogl bulin; therapeutic; streptokinase; de-immunised.  
 XX Streptococcus equisimilis.  
 XX WO9852976-A1.  
 XX 26-NOV-1998.  
 XX 21-MAY-1998; 98WO-GB01473.  
 XX 14-APR-1998; 98GB-0007751.  
 PR 21-MAY-1997; 97GB-0010480.  
 PR 31-JUL-1997; 97GB-0016197.  
 PR 28-NOV-1997; 97GB-0025270.  
 PR 02-DEC-1997; 97US-0067235.  
 XX (BIOV-) BIOVATION LTD.  
 PA Carr FJ;  
 PI WPI; 1999-045301/04.  
 DR Reducing immunogenicity of proteins - by modifying the amino acid  
 PT sequence of the protein to eliminate potential epitopes for T-cells  
 PT of a given species

Example 6; Fig 29; 77pp; English.

The invention relates to a method for the production of non-immunogenic proteins. The method comprises determining at least part of the amino acid sequence of the protein; (b) identifying in the amino acid sequence one or more potential epitopes for T-cells (T-cell epitopes) of the given species; and (c) modifying the amino acid sequence to eliminate at least one of the T-cell epitopes identified in step (b) thereby to eliminate or reduce the immunogenicity of the protein when exposed to the immune system of the given species. A method of analysing a pre-existing protein to predict the basis for immunogenic responses is also provided. The methods can be used particularly for reducing the immunogenicity of immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The products can be used for diagnosis and therapy. The present sequence represents the amino acid sequence of a de-immunised SK protein molecule.

Sequence 414 AA;

Query Match 6.5%; Score 91.5; DB 20; Length 414;  
 Best Local Similarity 25.3%; Pred. No. 0.68;  
 Matches 57; Conservative 31; Mismatches 86; Indels 51; Gaps 10;  
 QY 52 DDLQVIGASRILERSLSRSRVIANVLPFFHDPDIPGVTLTADPY-----VFQR 103  
 Db 95 ddyfevdfas---datitdrn-----gkvyfadkd-gsvtlptqpvgqefllghvrvr 144  
 QY 104 ELRLIYVGGDGPMSVGAIKAAVEVNPAGSFIVFSDAKADYHKNNELLQLLQK- 162  
 Db 145 pykekipnqa-----ksvdvetyvtqfplnpdddfrrglktilktaigt 193

QY 163 --SQVFLVLTGDCGDRTHPGYLAPEETAS--TSSGQVFLDQKQVSEVLKWFSAIQASK 218  
 Db 194 itsqellagqsilnkthpgytiyerdsslvthdnidfrtlpmdqeftyhvkneqaye 253  
 QY 219 VHLLSADHEE-----EGEHTWRIPFDPDS-LKEVTI 247  
 Db 254 inksglneeinntdlisekyvllkgekpyd-pfdrshklfti 297

RESULT 14

AAW01295  
 ID AAW01295 standard; protein; 414 AA.  
 XX  
 XX AAW01295;  
 XX 25-SEP-2000 (first entry)  
 XX Wild type streptokinase from Streptococcus equisimilis.  
 DE Immunogenicity; immunogen; T cell epitope; T-lymphocyte; drug;  
 KW vaccine; carrier; diagnosis; therapy; complement; C3; C5; C5F;  
 KW plasminogen; streptokinase; fibrin; blood clot; thrombolysis;  
 KW plasmin; myocardial infarction.  
 XX Streptococcus equisimilis.  
 OS WO200034317-A2.  
 XX 15-JUN-2000.  
 XX 08-DEC-1999; 99WO-GB04119.  
 XX 08-DEC-1998; 98GB-0026925.  
 PR 02-FEB-1999; 99GB-0002139.  
 XX (BIOV-) BIOVATION LTD.  
 PA Carr FJ, Adair FS, Hamilton AA, Carter G;  
 PI WPI; 2000-423372/36.

Rendering a protein non-immunogenic or less immunogenic useful in medicine and in diagnostics involves determining the amino acid sequence of the protein, identifying and modifying potential epitopes  
 Example 2; Fig 5; 42pp; English.

Proteins or their fragments can be rendered non-immunogenic or less immunogenic by identifying one or more potential T cell epitopes and modifying the sequence to eliminate at least one of these T cell epitopes to reduce the immunogenicity of the protein when exposed to the immune system of another organism. The method can be used for producing a protein with an enzymatic activity which has a beneficial therapeutic effect, a protein used to convert inactive drugs to its active form within a living organism, a vaccine, a protein used as a carrier of other molecule or a protein which binds to other molecules within or introduced within the living organism in order to alter the bio distribution of other molecules, such that the protein does not stimulate immune response in the living organism. The less immunogenic protein is useful in medicine, diagnosis and in manufacture of a therapeutic or a diagnostic agent. Streptokinase is produced by certain strains of beta-haemolytic streptococci. The protein has no inherent enzymatic activity but has considerable clinical importance owing to its ability to efficiently bind human plasminogen, potentiating its activation to plasmin and thereby promoting the dissolution of fibrin filaments in blood clots. Streptokinase is an effective thrombolytic agent in the treatment of coronary thrombosis, improving survival and preserving left ventricular function following myocardial infarction. The native protein is immunogenic and the production of neutralising antibodies in humans generally limits the protein to a single use. The new method could provide a longer therapeutic use for streptokinase. This is the sequence of the wild type streptokinase. The altered streptokinase sequence







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 21:07:27 ; Search time 36.48 seconds  
(without alignments)  
574.233 Million cell updates/sec

Title: US-09-665-728-1  
Perfect score: 1403  
Sequence: 1 MTPRAQLPLLLATYTVAA.....IEVRDPLGMSQSGSPPLLMQD 275

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_58:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	628.5	44.8	5175	2 T20992	hypothetical prote
2	628.5	44.8	5198	2 T43290	hemocentin precurs
3	110	7.8	2229	2 T16199	hypothetical prote
4	109	7.8	2014	2 T71560	hypothetical prote
5	106	7.6	2470	2 I50726	cation-independent
6	105.5	7.5	1363	2 C84346	hypothetical prote
7	105	7.5	476	2 T27707	hypothetical prote
8	105	7.5	2214	2 T16305	hypothetical prote
9	101.5	7.2	1324	2 T14070	peptide synthetase
10	96.5	6.9	1292	2 T48007	P-glycoprotein hom
11	96	6.8	895	2 I53908	major vault protei
12	93.5	6.7	1576	2 T28655	receptor-like hist
13	93.5	6.7	1576	2 T30898	receptor-like hist
14	93	6.6	458	2 T19941	hypothetical prote
15	93	6.6	1362	2 A75207	anypollulanase p
16	92.5	6.6	1387	2 A96771	hypothetical prote
17	92	6.6	666	2 T31461	probable magnesium
18	92	6.6	788	2 I51550	integrin beta-3 su
19	91.5	6.5	806	2 A46271	hypothetical prote
20	91	6.5	1250	2 T27706	RAD23 protein homo
21	90.5	6.5	392	2 T04150	protein-glutamine
22	90.5	6.5	693	2 B45991	protein disulfide
23	89.5	6.4	643	1 S32476	streptokinase G pr
24	88.5	6.3	440	2 S02723	parquat-inducible
25	88.5	6.3	547	2 F82162	lrp protein - huma
26	88	6.3	896	2 S57723	platelet-derived g
27	88	6.3	1098	1 PPM5RB	probable ABC trans
28	88	6.3	1229	2 F86155	P-glycoprotein-lik
29	88	6.3	1229	2 D85023	

30	88	6.3	1229	2 T52319	P-glycoprotein-lik
31	88	6.3	1278	2 E86155	probable ABC trans
32	87.5	6.2	552	2 E69438	probable fatty-aci
33	87	6.2	353	2 T45280	probable ABC trans
34	87	6.2	758	2 T02925	protoporphylin IX
35	86.5	6.2	440	1 A22801	streptokinase prec
36	86.5	6.2	1230	2 S47466	cellulose 1,4-beta
37	86	6.1	414	2 T30829	hypothetical prote
38	86	6.1	1914	2 T42635	tenascin Y precurs
39	85.5	6.1	754	2 T06249	protoporphylin IX
40	85.5	6.1	992	2 A42318	glycogen phosphory
41	85	6.1	300	2 T36768	secreted serine pr
42	85	6.1	415	1 B250	streptokinase (EC
43	85	6.1	575	2 S14955	hydroxymethylgluta
44	85	6.1	1081	2 T00330	hypothetical prote
45	85	6.1	1433	1 GNVUBW	M polyprotein prec

ALIGNMENTS

RESULT 1  
T20992  
hypothetical protein F15G9.4a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T20992; T24733  
R:Sulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20992  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <WIL>  
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a  
A:Experimental source: clone F15G9  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24733  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <WI2>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a  
A:Experimental source: clone T09B9  
C:Genetics:  
A:Map position: X  
A:Gene: CESP:F15G9.4a  
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3;  
A:Introns: 2512/2; 2593/3; 2699/3; 2759/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3;  
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1;

Query Match	44.8%	Score	628.5	DB	2	Length	5175
Best Local Similarity	48.6%	Pred. No.	5.6e+45				
Matches	126	Conservative	51	Mismatches	69	Indels	13
Gaps	3						
QY	10	LLLATYTVAAAVTSDEPTKTLSPATGDTATLAFVFDVTGSMWDDLMOVIDGASRLERSL	69				
Db	15	LLLAT---TCSSVNDK-----NDPTGKSLAFVFDITGSMFDLVQVREGAAKIFKTVM	66				
QY	70	SSRSRVIANVALVPHDPDIPVTLTADPVVFORLQLYVQGGDCPCPMSVGAIAKAAVE	129				
Db	67	AQREKLIVNIMVPHDPDIPVTLTADPVVFORLQLYVQGGDCPCPMTLTILKALQ	126				
QY	130	VANPGSIYVPSDARADYHKKNELLOLLOLQSQVVFVLTGDCGDRTHPGYLAFAFEIAS	189				
Db	127	ISLSPSIYVFTDARSXDYHLEDEVLNTIQEKSSVVFVMTGDCGNRTHPGFRTYKIAA	186				
QY	190	TSSQGVFOLDKQVSEVLKWNVESAIQASKVHLLSADHEEGEHTWRIPFDPKSEVTISL	249				
Db	187	ASFGQVHLEKSDVSTVLEYVRHAKVKKVHLMYEAREGGTVSRNPVDPKHSELTISL	246				



A;Residues: 1-2014 <WI3>  
A;Cross-references: EMBL:T25C12.3; PIDN:CAA91487.1; GSPDB:GN00028; CESP:T25C12.3  
A;Experimental source: clone T25C12  
C;Genetics:  
A;Gene: CESP:T25C12.3  
A;Map position: X  
A;Atrons: 10/2; 91/3; 136/1; 159/2; 182/3; 203/1; 225/3; 280/1; 422/3; 466/3; 559/3; 693/3; 1722/1; 1779/3; 1843/3; 1908/2; 1941/2; 2007/2

Query Match	7.8%	Score 109;	DB 2;
Best Local Similarity	23.0%;	Pred. No. 1.1;	
Matches 62;	Conservative 47;	Mismatches 117;	Indels 44;
Gaps 12;			

Qy		26	EPTKTLSPATGDATLAFVFDVTGSMWDDLMQVIDGASRIELERSLSR-----VIANY 79        : : : : :      :
Dd		1381	EPYPYPTFNDKDRTLVLMLTETSYNMGSSIFQL----KKNIKASLDISINDDTTQGWFNF 1436         : : : : :       :

**Qy** 80 ALVPF-----HDPDIGPVLTA---DPVVFRQLRYVQG---GSDCPEMSVGAKAAV 128  
||| : ||| : ||| : ||| : ||| : ||| :  
**Db** 1437 VLYPFDSITNKDSWYPPPTISRNDDIVTALKNISTMCPGAPCSSQCPRIVSVLSIL 1496

**Oy**    129 E---VANPGSFIYVFSARAKDYHKKNELQLLQKQSQQVFV----LTGDCGRD-THPGY 181  
       :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |  
**Dd**    1497 DMNALAAPNSVIWITRSSPEDYLQGQTISQKLTKKAYINFFPAIDSPCGEGWNSPNT 1556  
         :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

**Oy** 182 LAFEEIATSSGOVFQLDKQVSE--VLKWV-----ESAIQASKVHLLSADHEEGEHTW 234  
| | | | | : : : : | | : : : ||  
**Dd** 1557 DALFOITISYSGNGTFTMNAVDLSKNFLQYVPPTYLYSSGGTAASNGNCQTDEIIIFQVEH-- 1614

Qy	235	RIPFDSLKEVTISLSGPGPE-IEVRDPLG	263
		:   :     :   :	
Db	1615	-----EMYEFSIDFYHPLMETIKVFDPSG	1638

RESULT 5  
I50726

C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 20-Aug-1999  
C:Accession: I50726

Proc. Natl. Acad. Sci. U.S.A. 92, 9762-9766, 1995  
A>Title: Cloning and expression of the cDNA of chicken cation-independent mannose-6-phosphate 6-phosphotransferase  
A:Reference number: 150726, MUID:96003859  
A:Year: 1995

A: Status: preliminary; translated from GB/EMBL/DDBJ  
A: Molecule type: mRNA  
A: Residues: 1-2470 <ZHO>  
A: GC content: 68.9%  
A: GC content: 68.9%

C: Superfamily: mannose 6-phosphate receptor, cation-independent; fibronectin type II repeat F; 1832-1921/Domain: fibronectin type II repeat homology <2F1>

Query Match	7.6%;	Score 106;	DB 2;	Length 2470;
Best Local Similarity	23.4%;			
Matches	43;	Conservative	30;	Mismatches 59;
				Indels 52;
				Gaps 8;

QY 15 YTWAAATSD--EPTKLSPATGDTLAFVFDVTGSMDDLMQVDIGASILERSUSSR 72  
: | : | :  
Db 1478 FTAAACPLKSNVONDCRYMNPATG-----HLFDLTSLKRESGYTITDSSHNRKIELNYCAE 1532

Qy 73 SR-VIANVALVPFDPDIPGVTLTA---DPVVQRELRLQLVYQGSGDCPEMSVGAKAA 127  
:: || | | - | || | | : : : : | | |  
Db 1533 AKSSCANGAAVCTID---GPKTLNAGKLSKLTLYEDQVLKLVYEDGDPCP----- 1579

```
Qy   128 VEVANPGSFIYVFSARAKDYHKKNEQLLQLLKQSQQVFILTCGCDRTHPGYLAFEEI 187  
      :|: | ||        +||         |:| |:| |:| |:| |:| |:|  
Db   1580 -----TDLTKT--HK-----SYSFVKCKSDAGDDSQPVFLSFDEQ 1612
```

Qy 188 ASTS 191  
    ||  
Db 1613 TCST 1616

RESULT 6  
C84346  
hypothetical protein Vng1953c [Imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02  
C:Accession: C84346  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.;  
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.;  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt,  
A.:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483

A;Accession: C64340  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1363 <STO>

A:Accession: 1500  
A:Cross-references: GB:AE004437; NID:gl0581390; PIDN:AAG20135.1; GSPDB:GNO0138  
C:Genetics:  
A:Gene: VNG1953C

Query Match	7.5%	Score 105.5;	DB 2;	Length 1363;
Best Local Similarity	22.2%	Pred. NO. 1.3;		
Matches	58;	Conservative	33;	Mismatches 91;
				Indels 79;
				Gaps 10;

QY	13	ATYTVAAAVTSDEP-----TKTLPSPATGDTAFVFVDVTGSMWDDLQMVIDGASRI  64
	:      :      :      :      :      :      :      :      :      :      :      :	
Db	298	AASMAITADATTDPPLVVVPSTRTVNQSTDALDVAYRN-----QPVGSGSD  347

QY 65 LERSLS-----SRSRVIAN-----YAL-VPFHDPDIPV 92  
 : : | | : || | | | |  
Db 348 VSMTLAAANCTENATFDISDSPLVANATORLIDLAENAPLADGVYTSLVTVDGSGSPV 407

```

Qy 93 TLADPVVFQRELRLVQGDCPEMSVGAIKAAVEVANPGSFI--YVFSARAKDYH 149
    | | | | | : : : | : : | : : | : : | : : | : : | : : | : : |
Db 408 TRTDPVVVTD-----EPPDVSAVSLNASTDV-SPRDTVAVNVTYDDA----- 450

```

```

Qy 150 KKNE LLQLKQSQVVFVLTGDCGRTHPGYLAFEETASTSSGVQFQDKQVSEVLK 209
Db 451 -----TNATSATVHFVAADAGNFT-AGDLNAAASVSKDIDMAPCKQRDVNLRW 501

```

Qy 210 VESAIQASKVHLLSADHEEG 230  
| | : |  
Db 502 VREVTDSNYTVYVVTATDERG 522

## RESULT

hypothetical protein ZK193.3 - *Caenorhabditis elegans*  
C-species: *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27707  
R:Geisel, C.

submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of *C. elegans* cosmid ZK1193.  
A;Reference number: Z20409

A;reference number: 420403  
A;Accession: T27707

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA

A;Molecule type: DNA  
A;Residues: 1-476 <GEI>

A; Cross-references: EMBL:U41553; PIDN:AAA83292.1; CESP:2K1193.3  
G. Genetics:

C;Genetics:  
A;Gene: CESP:ZK1193.3

A; Introns: 22/3; 132/1; 182/3; 308/1; 358/3; 403/1

Query Match	Score 105;	DB 2;	Length 476;
7.5%			

Best Local Similarity 22.3%; Pred. No. 0.31;  
Matches 53: Conservative 49: Mismatches 100: Indels 36: Gaps 10:

Query Match	7.2%;	Score 101.5;	DB 2;	Length 1324;
Best Local Similarity	20.6%;	Pred. No. 2.7;		
Matches 67:	Conservative	38:	Mismatches 123.	Indels 07.
				Caps 11

QY	133	PGSFIYVFS	DARAKYH	KKNELLL	QLQLKQ	SQVVFVLT	DGCDRTHP	PGYLAFEE	IASTSS	132
DB	278	TGALLAAFP	PAPPTDA	ERAGPL	-----	PGHLAYL	LLHTSGS	314		
QY	193	GOV--	FOLDKQV	SEVLKVS	IAQKVL	LSADHEE	EGEHTWR	PPFOPSKE	VITSLS	250
DB	315	GRPKGVAV	GHQVAT	LLSAGT	CVGADR	LHRTVAT	SS-----	FQVSVFD	TLVPL	366

QY 251 GPGPEIEVRDPLGMSQ--GSPPLL 272  
| | : | : | | |  
Db 367 TGGRIEIVENTLAVADRTGGEPSLL 391

RESULT 10  
T48007  
P-glycoprotein homolog Tl7Jl3.ll0 [similarity] - Arabidopsis thaliana  
N-Alternate names: protein Tl7Jl3.ll0  
C-Species: Arabidopsis thaliana (mouse-ear cress)  
C-Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 18-Aug-2000  
C-Accession: T48007  
R-R; Riger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Lemcke  
A-Submitted to the Protein Sequence Database, February 2000  
A-Reference number: 224482  
A-Accession: T48007  
A-Status: preliminary  
A-Molecule type: DNA  
A-Residues: 1-1292 <RIE>  
A-Cross-references: EMBL:AL138651  
A-Experimental source: Cultivar Columbia; BAC clone Tl7Jl3  
A-Map position: 3

[illegible]

A;Cross-references: EMBL:U09870; NID:q665922; PIDN:AAC52161.1; PID:q497940

Query Match  
6.7%; Score 93.5; DB 2; Length 1576;

RESULT 14  
T19941  
hypothetical protein C44H4.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T19941  
R:Smeye, R.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19200  
A:Accession: T19941  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-458 <WIL>  
A:Cross-references: EMBL:Z79598; PIDN:CAB01867.1; GSPDB:GN00028; CESP:C44H4  
A:Experimental source: clone C44H4  
C:Genetics:  
A:Gene: CESP:C44H4.1  
A:Map position: X  
A:Introns: 24/3; 157/1; 257/3; 372/3; 425/1

```

Query Match      6.6%; Score 93; DB 2; Length 458;
Best Local Similarity 19.4%; Pred. No. 3.2;
Matches 56; Conservative 51; Mismatches 99; Indels 82; Gaps 14;

QY 11 LLATYVVAAVTSDPTKTL-SPATGDATLAFVFDVT---GSMDDLLMQVIDGASRILE 66
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 3  LLLFNLLIAGWVYSCQPTLQLQEPCTCTSTRYEAVSINCDSGLDAVLESLSNSPQAI 62
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 67 RLSLSRSRV-----IANYALVPFHPDIPGVT 93
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 63 SLTISNTPIEKVLVFNHDKISLYLFEOMPQYAFQGGFTKKLFRLNGLRSPH-----PNT 117
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 94 LTADPVVYFQRELRLYVQGG--GDCPEMSYGAIAAAVEAVANPGSFIYVFSARAKDVHK 151
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 118 FTGN---LENSLEELRGNVIGDIPGSGVSILKQLKILSPDLNIEYVODNAFLSYHSR 174
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 152 NELLQLQLKOSQVVFVLTGDCDRTHP-GYLAFAEEIASTSSGVOFOLDKQOVSEV---- 206
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 175 DSLUK-LDLSANNLTAI-----HPTGLLGLLENLSQLS-----LDKNLLSEIP 218
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 207 LKWVES---AIQASKVHLLSADHEEGEHTWPIFPDPSLKEVNTISUS 250
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 219 LENIPSLDLSLGNWRTHTLSRN-----SIPL-PNEKLSLSLEVY 256
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

RESULT 15

A75207  
amylotopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: A75207  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: A75207  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1362 <RAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49104.1; PID:e151499  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: apu; PAB0122

	Query Match	6.6%	Score 93;	DB 2;	Length 1362;
	Best Local Similarity	21.5%;	Pred. No. 15;		
	Matches 65;	Conservative 42;	Mismatches 105;	Indels 90;	Gaps 14;
QY	17	VVAAAVTSDPTK	-----TLSP-ATGDATAFVEDVTGSMWDDLMOQVDGASRIERSLSS	71	
		:	:	:	
Db	622	IVRALVGLPEGVKNNSSLSPLAG	-----IEVYFD-----DEGLHFVVLNTRSFSEIYE	672	
		:	:	:	
QY	72	RSRVIAN	-----YALVPFHDPDGPVTLTADPV-----VFQELRQLXVQG	112	
		:	:	:	

```

Db 673 PEKIIGNTFTVLQKKPEEFRYSEVPFSDKSVGLLITHTITVKGGEVFVKATSYDNYKKV 7332
QY 113 GGDCEPMSVGAIKAAVEVANPGSFI-----YVFSDDAKVDYHKKNELLQLQLK-----Q 1622
Db 733 G-----EVKYNAINGYEVVVPFDYIEPDSFYFAVSTINDNGSLSEIITTPIHUKLPKEIE 788
QY 163 SQVFLVLTGDCDTRTHGYLAFEEIASTSSGOVFLQDKQVSEVLKWSAIOASQVHLL 222
Db 789 GTLITEIKDIEGDDHGPENVTY-----ATDKVF-----VEHHLLDLKVRLL 829
QY 223 SADH:-----BEEGETHWIRPDPSLKEVTLSL-----SGGPEIEVR 259
Db 830 ERPSNVFEFYFKELGDNPNWAPYGLSQIMEVYLDYKEGNTSAIKMFPDGPGSNDL- 888
QY 260 DP 261
Db 889 DP 890

Search completed: March 7, 2002, 22:06:20
Job time: 3533 sec

```

Search completed: March 7, 2002, 22:06:20  
Job time: 3533 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:05:27 ; Search time 26.04 Seconds  
(without alignments)  
387.205 Million cell updates/sec

Title: US-09-665-728-1  
Perfect score: 1403  
Sequence: 1 MTPRAQLPLLLATYTVVAA.....IEVRDPLGMSQSPPLMQD 275

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	6.8	885	1 MVP_RAT	Q62667 rattus norv
2	93.5	6.7	885	1 CHLD_SYNP7	O07345 synechococc
3	92.5	6.6	1387	1 PUR4_ARATH	Q9m8d3 arabidopsis
4	92	6.6	666	1 BCHO_HELMO	O3zge6 heliobacill
5	91.5	6.5	806	1 ITB7_MOUSE	P26011 mus musculu
6	90.5	6.5	693	1 TGL3_MOUSE	Q08189 mus musculu
7	89.5	6.4	643	1 PDA4_RAT	P38659 rattus norv
8	88.5	6.3	440	1 STRP_STRSP	P10519 streptococc
9	88.5	6.3	515	1 PDL_ASPOR	Q00248 aspergillus
10	88	6.3	896	1 MVP_HUMAN	Q14764 homo sapien
11	88	6.3	1098	1 PGDR_MOUSE	P05622 mus musculu
12	87.5	6.2	158	1 GRE4_RHILV	O88546 rhizobium l
13	87	6.2	758	1 CHLD_TOBAC	O24133 nicotiana t
14	86.5	6.2	440	1 STRP_STREQ	P00779 streptococc
15	86	6.1	403	1 PTEN_HUMAN	O00633 homo sapien
16	86	6.1	403	1 PTEN_MOUSE	O08586 mus musculu
17	85.5	6.1	754	1 CHLD_PEA	O22437 pisum sativ
18	85.5	6.1	992	1 PHS2_DICDI	P34114 dictyostell
19	85	6.1	575	1 HMD1_HEYBR	P29057 hevea bras
20	85	6.1	1433	1 VGLM_BUNYW	P04505 bunyamwera
21	84.5	6.0	2128	1 SPCB_MOUSE	P15508 mus musculu
22	84	6.0	759	1 CHLD_ARATH	O9sje1 arabidopsis
23	83.5	6.0	695	1 ANPH_HUMAN	P49418 homo sapien
24	83	5.9	507	1 ATPA_CHLRE	P26526 chlamydomon
25	82.5	5.9	158	1 GRE4_RHINE	P36894 rhizobium m
26	82.5	5.9	344	1 YK58_ARCFU	O28221 archaeglob
27	82.5	5.9	1260	1 YAOE_SCHPO	Q10093 schizosacch
28	82.5	5.9	2137	1 SPCB_HUMAN	P12777 homo sapien
29	82.5	5.9	3649	1 ACVS_NOCLA	P27743 nocardia la
30	82	5.8	506	1 ATPA_CHLVU	P56294 chlorella v
31	81.5	5.8	412	1 AK_PSEAE	O69077 pseudomonas
32	81.5	5.8	913	1 CAD4_MOUSE	P39038 mus musculu
33	81.5	5.8	1106	1 PGDR_HUMAN	P09619 homo sapien

34 81 5.8 411 1 YG12\_METJA  
35 81 5.8 788 1 ITB3\_HUMAN  
36 81 5.8 920 1 PARC\_SYNY3  
37 80.5 5.7 424 1 THIK\_HUMAN  
38 80.5 5.7 455 1 SYS\_PYRHO  
39 80.5 5.7 487 1 XYL\_C\_PSEPU  
40 80.5 5.7 714 1 PBPF\_BACSU  
41 80.5 5.7 798 1 ITB5\_MOUSE  
42 80.5 5.7 838 1 OS9A\_MOUSE  
43 80.5 5.7 916 1 MUPS\_RHIME  
44 80.5 5.7 1257 1 PGCN\_RAT  
45 80 5.7 476 1 REAE\_HAEIN

Q59007 methanococc  
P05106 homo sapien  
P73077 synechocyst  
P09110 homo sapien  
O58441 pyrococcus  
P43503 pseudomonas  
P38050 bacillus su  
O70309 mus musculu  
P48722 mus musculu  
P56883 rhizobium m  
P55067 rattus norv  
O05074 haemophilus

## ALIGNMENTS

RESULT 1  
MVP\_RAT  
ID MVP\_RAT STANDARD; PRT; 885 AA.  
AC Q62667;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE MAJOR VAULT PROTEIN (MVP).  
GN MVP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95129871; PubMed=7828886;  
RA "Kickhoefer V.A., Rome L.H.;  
RT "The sequence of a cDNA encoding the major vault protein from Rattus  
RT norvegicus.";  
RL Gene 151:257-260(1994).  
CC -!- FUNCTION: UNKNOWN, THOUGH MVP IS REQUIRED FOR NORMAL VAULT  
CC STRUCTURE. VAULTS ARE MULTI-SUBUNIT STRUCTURES THAT MAY BE  
CC INVOLVED IN NUCLEO-CYTOPLASMIC TRANSPORT.  
CC -!- SUBUNIT: MAJOR COMPONENT OF THE VAULT, A LARGE RIBONUCLEOPROTEIN  
CC PARTICLE. THE VAULT CONTAINS FOUR PROTEINS OF MW 210 KDA, 192 KDA,  
CC 104 KDA AND 54 KDA, RESPECTIVELY, AND A SPECIES OF SMALL RNA OF  
CC CIRCA 140 BASES IN RELATIVE MOLAR RATIOS OF 3:1:55:7:9.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: STRONG, TO D.DISCOIDEUM MVP AND MVPB.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U09870; AAC52161.1; ALT\_INIT.  
DR InterPro: IPR002499; Vault.  
DR Pfam: PF01505; Vault; 7.  
KW Ribonucleoprotein.  
SQ SEQUENCE 885 AA; 98504 MW; EA7F61A0AD365013 CRC64;

Query Match 6.8%; Score 96; DB 1; Length 885;  
Best Local Similarity 21.7%; Pred. No. 2.8;  
Matches 65; Conservative 43; Mismatches 101; Indels 90; Gaps 16;  
QY 4 RAQLPLLLAT-----YTVVAAVTSDEPTKTLSPATGATATLAFVDFVTGSMWMDLMQVI 58  
Db 42 RVLPAPVMVTPPHYCVANPVSRDQTQSSVLFDTQVRLRHA-DQEIRLAQDPFFLY 100  
QY 59 DGASRLRSLSRSRVIAN-----YALVPHPDDPGVPTLTADPVVQRELRLQLYVGG 113  
Db 101 PG--EVLEKDTITPQVLPNTALHLKLLDFEDKN-GDKVMAGDEWLFE----- 146

QY 114 GDCPEMSYGAIAKAAEVANPGSFYIVSFDSARAKDYHKKNELLOLLOLQ---LKOSQVVFV-L 169  
 Db 147 -----GPGTYI-----PQREVEVEIIQATVIKQNALRLRA 178  
 QY 170 TGDGCDRTHPGYLAPEEIASTSSG---QVFQDKQVSEVYLKWKVESAIQASK--VHLLS 223  
 Db 179 RKECFDREGKRVGTGEEMVRSVGYLPAVFE-----EVLDLVDVAIVLTETALHLRA 231  
 QY 224 ADH-----EEEGHTW-----RIP--FDPSLKEVTISLGGPGEIEVRDPLG 263  
 Db 232 LQNFRLRGVLRHTGEE--WLVTVDTEAHVPDYVEEVLGVVPIITLTPRHVCYILDPMG 289  
 RESULT 2  
 CHLD\_SYN7  
 ID CHLD\_SYN7 STANDARD; PRT; 677 AA.  
 AC O07345;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAGNESIUM-CHELATASE SUBUNIT CHLD (MG-PROTOPORPHYRIN IX CHELATASE) (MG-  
 CHELATASE SUBUNIT D).  
 GN CHLD.  
 OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 OX NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Masuda T., Nakayama M., Ohta H., Takayama K.-I.;  
 RT "Cloning and sequencing of a chld gene encoding a subunit of  
 magnesium-chelatase from the cyanobacterium Synecococcus sp. PCC  
 7942.";  
 RL (In) Plant Gene Register PGR97-091.  
 CC -I- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES  
 A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-  
 PROTOPORPHYRIN IX.  
 CC -I- PATHWAY: CHLOROXYLL BIOSYNTHESIS.  
 CC -I- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.  
 CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB0031135; BAA20346.1;  
 DR InterPro: IPR000523; MG\_chelatase\_chiI.  
 DR Pfam: PF01078; Mg\_chelatase; 1.  
 DR SMART: SM00327; VWFA; 1.  
 DR PROSITE: PS50234; VWFA; 1.  
 KW Photosynthesis; Chlorophyll biosynthesis.  
 FT DOMAIN 481 676 VWFA.  
 FT DOMAIN 329 383 GLU/PRO-RICH.  
 SQ SEQUENCE 677 AA; 73367 MW; 7E2988E096DF5506 CRC64;

Query Match 6.78; Score 93.5; DB 1; Length 677;  
 Best Local Similarity 24.18; Pred. No. 3.2;  
 Matches 52; Conservative

QY 24 SDEPTKTLSPATGATIAFVDVTGSGMMDLMOVIDCA-SKILERSLSRSRVNTANALV 82  
 Db 467 ADIRSKQLVRKAG-ALVIFLVDASGSMALNMQSAKGAIVRLTTEAYENRDQV---ALI 521  
 QY 83 PFHPDIDGPVTL--TADPVVVFQRELRLQLYVQGGDCPEMSYGAIAKAAEVANPGSF--- 136  
 Db 522 PFRG-EQAEVLLPPTRSITTAARKRLKMKPCGGGLPLAHLGTQAVRVCTNAQSGDIGQVV 580

QY 137 IYVESDARA-----KOYHK---KNELLQLLQKQSVVFLTDCGDRTHPGY 181  
 Db 581 IVALTDGRGNIPLARSIGQPMEEGKPDLEELDLIAKRIGLSKQLLYDIT-PRKRVGA 639  
 QY 182 LAPEEIASTSSGQVFQDKQVSEVYLKWKVESAIQAS 217  
 Db 640 GFGEKLANAAGGQYHLPKVSQDAIAAAMAQSAALRAT 675  
 RESULT 3  
 PURA\_ARATH  
 ID PURA\_ARATH STANDARD; PRT; 1387 AA.  
 AC Q9M8D3;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE PROBABLE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE, CHLOROPLAST  
 DE PRECURSOR (EC 6.3.5.3) (FGAM SYNTHASE) (FGAMS) (FORMYLGLYCINAMIDE  
 DE RIBOTIDE AMIDOTRANSFERASE) (FGARAT) (FORMYLGLYCINAMIDE RIBOTIDE  
 DE SYNTHETASE).  
 GN ATIG74260 OR FIO17.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=2016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:816-820(2000).  
 CC -I- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +  
 L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-  
 FORMYLGLYCINAMIDINE + L-GLUTAMATE.  
 CC -I- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.  
 CC -I- SUBCELLULAR LOCATION: CHLOROPLAST (PROBABLE)  
 CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS  
 CC FAMILY  
 CC -I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO TYPE-1 GLUTAMINE  
 CC AMIDOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AC020579; AAG52403.1;  
 DR InterPro: IPR000728; AIRS\_related.  
 DR Pfam: PF00586; AIRS; 1.  
 KW Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase;  
 KW Chloroplast; Transit peptide.  
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).

[illegible]



DE (TGASE E3) (TRANSGLUTAMINASE 3).  
GN TGM3 OR TGASE3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/C; TISSUE=Epidermis;  
RX MEDLINE=93286109; PubMed=8099584;  
RA Kim I.-G., Gorman J.J., Park S.-C., Chung S.-I., Steinert P.M.;  
RT "The deduced sequence of the novel protransglutaminase E (Tgase3) of human and mouse."  
RL J. Biol. Chem. 268:12682-12690(1993).  
CC -!- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE CONJUGATION OF POLYAMINES TO PROTEINS. IT IS RESPONSIBLE FOR THE LATER STAGES OF CELL ENVELOPE FORMATION IN THE EPIDERMIS AND THE HAIR FOLLICLE.  
CC -!- CATALYTIC ACTIVITY: PROTEIN GLUTAMINE + ALKYLAMINE = PROTEIN N(5)-ALKYLGLUTAMINE + NH(3).  
CC -!- COFACTOR: CALCIUM IS REQUIRED FOR TRANSGLUTAMINASE ACTIVITY.  
CC -!- SUBUNIT: THIS ENZYME CONSISTS OF TWO POLYPEPTIDE CHAINS, WHICH ARE SYNTHESIZED AS A PRECURSOR FORM OF A SINGLE POLYPEPTIDE.  
CC -!- SIMILARITY: BELONGS TO THE TRANSGLUTAMINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L10385; AAA40421.1; .  
DR HSP: P00488; 1GCT.  
DR MGD: MGI-98732; Tgm3.  
DR InterPro; IPR002931; Transglut\_core.  
DR InterPro; IPR001102; Transglutmmse.  
DR Pfam; PF00927; Transglutamin-C; 1.  
DR Pfam; PF00868; Transglutamin-N; 1.  
DR Pfam; PF01841; Transglut\_core; 1.  
DR SMART; SM00460; TGC; 1.  
DR PROSITE; PS00547; TRANSGLUTAMINASES; 1.  
DR Transferrase; Acyltransferase; Calcium-binding; zymogen.  
KW CHAIN 1 470 50 KDA NONCATALYTIC CHAIN.  
FT CHAIN 471 693 27 KDA CATALYTIC CHAIN.  
FT ACT\_SITE 273 273 BY SIMILARITY.  
SQ SEQUENCE 693 AA; 77324 MW; 8B3F0601735D4E84 CRC64;  
  
Query Match 6.5%; Score 90.5; DB 1; Length 693;  
Best Local Similarity 21.8%; Pred. No. 5,9;  
Matches 60; Conservative 29; Mismatches 71; Indels 115; Gaps 14;  
  
QY 48 GSWDDLMQVIGASRILERSLSRSRVANY----- 79  
Db 271 GOCV-----VFAGTLNVLRCGLGVRSVITFNESAHTDRNLSDVYVYDAMGNPLEKGS 325  
QY 80 ALVPEH-----DPDTPGV-----TLTADPVVQRELRLQLYVGGDCCPMSVGA 126  
Db 326 SWNFHVNNEGFWRTDUGPTTNGVQVLDATP-----QBSQGVFCQGPASVNAIKA 377  
QY 127 AVEANPGSFIYVFSARAKD-----YHKKNELLQLLQKSOQVFWLTDGCDGRTHP--G 180  
Db 378 G-DVDRNFDIMFIAEVNADRTIWNRRN-----TQKQNSV-----DRHSIGK 421  
QY 181 YLAFIEIASTSSGQVQLDKQVSEVLKWKVESIAQSKVHLLSAD-----HE 227  
Db 422 YISTRVAGSNS-----RMDVTDKIKYPEGSSEERVOHQALDKLPNASFGATSSRN 473  
QY 228 EGEHTWRIPFPDPSL-----KEVTISL 249  
Db 474 PEGEDK-----EPSSIGKFVKVTGILAVKGVEVSL 503

RESULT 7  
PDA4\_RAT PDA4\_RAT STANDARD; PRT; 643 AA.  
AC P38659;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROTEIN DISULFIDE ISOMERASE A4 PRECURSOR (EC 5.3.4.1) (PROTEIN ERP-72) (ERP72) (CALCIUM-BINDING PROTEIN 2) (CABP2).  
GN PDIA4 OR CABP2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar; TISSUE=Liver;  
RX MEDLINE=93238767; PubMed=8477750;  
RA Van P.N., Rupp K., Lampen A., Soeling H.-D.;  
RT "CABP2 is a rat homolog of ERP72 with protein disulfide isomerase activity."  
RL Eur. J. Biochem. 213:789-795(1993).  
CC -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.  
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.  
CC -!- INDUCTION: UPON GLUCOSE STARVATION, AS WELL AS TREATMENT WITH TUNICAMYCIN.  
CC -!- PTM: O-GLYCOSYLATED.  
CC -!- SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M86870; AAA19217.1; .  
DR HSP: P07237; IMEX.  
DR MEROPS; C17.002; .  
DR InterPro; IPR000886; ER target.  
DR InterPro; IPR000063; Thioered.  
DR Pfam; PF00085; thioered; 3.  
DR PRINTS; PR00421; THIOREDODXIN.  
DR PROSITE; PS00194; ER\_TARGET; 1.  
DR PROSITE; PS00194; THIOREDODXIN; 3.  
DR Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal;  
KW Glycoprotein; Calcium-binding  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 643 PROTEIN DISULFIDE ISOMERASE A4.  
FT DISULFID 89 92 REDOX-ACTIVE (BY SIMILARITY).  
FT DISULFID 204 207 REDOX-ACTIVE (BY SIMILARITY).  
FT DISULFID 553 556 REDOX-ACTIVE (BY SIMILARITY).  
FT SITE 640 643 PREVENT SECRETION FROM ER.  
SQ SEQUENCE 643 AA; 72806 MW; 85611B66F3A4DE41 CRC64;  
  
Query Match 6.4%; Score 89.5; DB 1; Length 643;  
Best Local Similarity 22.0%; Pred. No. 6,5;  
Matches 53; Conservative 35; Mismatches 76; Indels 77; Gaps 11;  
  
QY 15 YTVVAAVTSDEPTKLS--PATGATLAFVDFVTGSMWDDL-----QVIDGASRILERSL 69  
Db 99 YEKIASTLKNDPPIAVAKIDATSASMLSKFDVSGYPTIKLKKGQAVDYDGSRTQBEI 158  
QY 70 SRSRVIANALVPHDPDGP---VTLTA-----DPVVFQBE--LRQLYVGGGDCPEM 119  
Db 159 VAKVREVS-----QPDWTPPPPEVTTLTKENFDVYVNNADIILVEFYAPWCGHCKKL 210  
QY 120 SVGAIKAAVEVANPG-----SFIYVFSARAKDYHKKNEL 154

```
Db 211 APEYKAAKELSRPPIPLAKVDATQTLAKRFDVSGYPTLKIFRKRGRPDYNGPREK 270
QY 155 L-----QLLQKQSQVVFVLTGD-----CGDRTHPGYLAFEEETAS 189
Db 271 YGIVDMYVQSGPPSKEITLTKQVO-EFLKGDGDDVVILGVFGVGVD---PGYLQYODAA 326
QY 190 T 190
Db 327 T 327
```

```
RESULT 8
STRP_STRP STANDARD; PRT; 440 AA.
AC P10519;
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE STREPTOKINASE G PRECURSOR.
GN SKG.
OS Streptococcus sp. (strain 19909).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89160265; PubMed=2922269;
RA Walter F., Siegel M., Walke H.;
RT "Nucleotide sequence of the streptokinase gene from a group-G
RT Streptococcus.";
RL Nucleic Acids Res. 17:1262-1262(1989).
CC -!- FUNCTION: THIS PROTEIN IS NOT A PROTEASE, BUT IT ACTIVATES
CC PLASMINOGEN BY COMPLEXING WITH IT. AS POTENTIAL VIRULENCE FACTOR,
CC IT IS THOUGHT TO PREVENT THE FORMATION OF EFFECTIVE FIBRIN
CC BARRIERS AROUND THE SITE OF INFECTION, THEREBY CONTRIBUTING TO THE
CC INVASIVENESS OF THE CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13400; CAA31766.1; -.
DR PIR; S02723; S02723.
KW Plasminogen activation; Signal; Virulence.
FT SIGNAL 1 26
FT CHAIN 27 440 STREPTOKINASE G.
SQ SEQUENCE 440 AA; 50199 MW; 5521F8825FE1B6EA CRC64;
```

```
Query Match 6.3%; Score 88.5; DB 1; Length 440;
Best Local Similarity 25.3%; Pred. No. 4.7;
Matches 57; Conservative 30; Mismatches 87; Indels 51; Gaps 10;

QY 52 DDLQVIDGASRLERSLSRSRVIANALVPFHPDPDIPGVTLTADPV-----VFQR 103
Db 121 DDFEVIDFAS---DATITDRN-----CKVVFADKD-GSVTLPIQPVQEFLLKGHVRV 170
QY 104 ELRLQYVQGGDCPMSVGAIAKAAVEVANGPSFIYVFSARAKYHKKNELLQLQLK 162
Db 171 PYKEKPVQNOA-----KSDVDEYTVQFTPLNPDDDFRPALKDKTKLEKTLAIGT 219
QY 163 --SOVVELLTGCGDRTHPGYLAFEEIAS--TSSGQVFDLQKQVSEVLKWKVESATQASK 218
Db 220 ITSQELAAQASILKNHHPGTYIYERDSSIVTHDNDIFRTILPMDQEPYTHVKNREQAYR 279
QY 219 VHLLSADHEE-----EGEHTWRIPDPS--LKEVTI 247
Db 280 INKSKSLNEEINNTDLISEKYYVLLKKGKPYD-PFDRSHLKLFTI 323
```

```
RESULT 9
PDL_ASPO
ID PDL_ASPO STANDARD; PRT; 515 AA.
AC Q00248;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTEIN DISULFIDE ISOMERASE PRECURSOR (EC 5.3.4.1) (PDI).
GN PDI.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIB 40;
RA Lee B., Yamada O., Kitamoto K., Takahashi K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D85900; BAA12913.1; -.
DR HSSP; P07237; IMBK.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; ThioRed; 3.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 515 PROTEIN DISULFIDE ISOMERASE.
FT DISULFID 58 61 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 393 396 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 512 515 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 515 AA; 56458 MW; A155B3B8E20E37EAC CRC64;
```

```
Query Match 6.3%; Score 88.5; DB 1; Length 515;
Best Local Similarity 18.5%; Pred. No. 5.8;
Matches 63; Conservative 52; Mismatches 109; Indels 117; Gaps 15;

QY 7 LLPLLLATYVVAAVTSDEPTKTLSPATGATLAFVDFVTGSMWDDL----- 55
Db 8 ILSLLGASAVASAAATAEAPSDDVVS-LTGDTFTFVKE-----HDLVLAEFFAPWCGH 60
QY 56 -----QVTDGASRLERSLS-----SRSRVIANY-----ALVPFHD- 86
Db 61 CKALAPKYEQAAATELKEKNIPLVKVDCTEEALCRDQGVGYPTLKIFGLDAVPYQGA 120
QY 87 -----PDIGPVT-----LTADPV-----FQREL 106
Db 121 RQTEAIVSVWVKQSLPAVSPVTPENLEEIKTMKIVVICIASDDQANDIFTTAESOR 180
QY 107 QLYVQGGDCPMSVGAIAKAAVEVANGPSFIYVFSARAKYH---KNELLQLQLKQS 163
Db 181 DNYLFAA-----TSDASTAKAEGVKQPSIVLYKDFDENKATYDGETEQDALLSWKVTAST 235
```

**Qy**      4    RAQLPLLLAT-----YTVVAAAVTSDEPTKTLSPATGDTLAFVFDVTGS----- 49  
         |     :     :     |     |     :     :     :     :     :     :     :  
**Db**      42   RVLFAPRMWTVTPRRHYCTVA-----NPYSRDAQGLVLFDVTVGVFLRHADLE 89  
  
**Qy**      50   -MMDDLQMVIDGASRLERSLSRSRVIAN-----YALVPFHDPDIPGTLTADPWFQ 102



```
DR Pfam; PF00047; ig; 4.
DR Pfam; PF00069; pkinase; 2.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00410; IG_Like; 2.
DR SMART; SM00219; Ty_KC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Tyrosine-protein kinase: Receptor; Transmembrane; Glycoprotein;
KW Transferase; Phosphorylation; ATP-binding; Immunoglobulin domain;
KW Signal.
FT SIGNAL.
FT CHAIN 1 31
FT CHAIN 32 1098
FT
FT BETA PLATELET-DERIVED GROWTH FACTOR
FT RECEPTOR.
FT DOMAIN 32 530
FT TRANSMEM 531 555
FT POTENTIAL.
FT DOMAIN 556 1098
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 119
FT IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 128 209
FT IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 214 310
FT IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 415 527
FT IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 599 613
FT PROTEIN KINASE.
FT NP_BIND 605 631
FT BINDING 632 633
FT ACT_SITE 825 825
FT DISULFID 53 99
FT DISULFID 148 189
FT DISULFID 234 290
FT DISULFID 435 507
FT MOD_RES 750 750
FT MOD_RES 856 856
FT CARBOHYD 44 44
FT CARBOHYD 88 88
FT CARBOHYD 102 102
FT CARBOHYD 214 214
FT CARBOHYD 291 291
FT CARBOHYD 306 306
FT CARBOHYD 353 353
FT CARBOHYD 370 370
FT CARBOHYD 444 444
FT CARBOHYD 467 467
FT CARBOHYD 478 478
FT SEQUENCE 1098 AA; 122805 MW; 8D391CAFAC3FC31D CRC64;
Query Match 6.3%; Score 88; DB 1; Length 1098;
Best Local Similarity 18.7%; Pred. No. 18;
Matches 56; Conservative 45; Mismatches 107; Indels 92; Gaps 12;
QY 20 AAVTSDEPTKTLSPATGDAITLAFVFDVTGSMWDDLMQVIDGASRLERSLSRSRVIANV 79
DB 788 ATLINDSPVLSYTDLVG-----FSYQVANGM--DFL-----ASKNCVHRDLAARNVLICEG 836
QY 80 ALVPEHPDPDG-----PV-----TLTADPVVVFQRELRLQLY 109
DB 837 KLVKICDFGLARDIMRDSNYSISKSTYLPKWWAPESIFNSLYTTLSDVWSFGILLMEIF 896
QY 110 VQGGDCGPEMSV-----GAIKAAVEVANPGSFYVFDARAKDYHKKNELQLQLKQSQ 164
DB 897 TLGGTPPELPMNDQFYNAIKRGVMAQPA-----HASDEYEINQ----- 937
QY 165 VVFLTGCGDQ---RTHPGYLAFFEIASTSSGOVFLDKQOVSEVLKWWESAIAQSKVHL 221
DB 938 -----KCWEKEPFRPPESQLVLLERLLGEGYKKYQQVD-----EELFRSDHPAI 984
QY 222 LSADHEEEGETWRIPD-----PSLKEVTISLSGPGPIEVDRDPLGMSQSGPPL 271
DB 985 LRQAPPGIHSURSPDLDTSSVLYTAVQPNESDNDYTIPLDPKPDVADE-GLPEGSPSL 1043
RESULT 12
GREA_RHILV
```

```
ID GREA_RHILV STANDARD; PRT; 158 AA.
AC O68546;
DC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION ELONGATION FACTOR GRE A (TRANSCRIPT CLEAVAGE FACTOR
DE GRE A).
GN GREA.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3855;
RC MEDLINE=98434549; PubMed=9756877;
RA Kadmas J.L., Allaway D.; Studholme R.E., Sullivan J.T., Ronson C.W.,
RA Poole P.S., Raetz C.R.H.;
RT "Cloning and overexpression of glycosyltransferases that generate the
RT lipopolysaccharide core of Rhizobium leguminosarum.";
RL J. Biol. Chem. 273:26432-26440(1998).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GRE A OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3'TERMINUS. GRE A RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GRE A/GREB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF050103; AAC05214.1; -.
CC HSP; P21346; 1GRJ.
CC InterPro; IPR001437; GreA_Greb.
CC Pfam; PF01272; GreA_Greb; 1.
CC ProDom; PD004918; GreA_Greb; 1.
CC PROSITE; PS00829; GREAB_1; 1.
CC PROSITE; PS00830; GREAB_2; 1.
CC Transcription regulation; DNA-binding; Coiled coil.
KW SEQUENCE 158 AA; 17444 MW; FACA215D8CAF1605 CRC64;
SQ
```

```
Query Match 6.2%; Score 87.5; DB 1; Length 158;
Best Local Similarity 24.3%; Pred. No. 1.4;
Matches 36; Conservative 30; Mismatches 55; Indels 27; Gaps 7;
QY 133 PGSIYVFSARAKDYHKKNELQLQLKQSQVFLVLTGDCGDRTHPGYLAFFEIASTSS 192
DB 9 PGGEVFKQEELRWQRQEPRIIEAIAEAAH-----GDLSENAE---YHAAKEAQSHNE 60
QY 193 GOVFOLD-----KQOVSEVLKWWESAIAQ-ASKVHLLSADHEEGETWRIPD--PSLKEV 245
DB 61 GRISLEDLTARAEVLDITKMSGDKIRKPAKLIDEETEE--EKTQIVGQDEADVAKG 118
QY 246 TISLSGP-----GPEIEVRDPLG 263
DB 119 RISISSPTARALICKEVGDSEIENAPGG 146
RESULT 13
ID CHLD_TOBAC STANDARD; PRT; 758 AA.
AC O24133;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
ID GREA_RHILV
```



20-AUG-2001 (Rel. 40, Last annotation update)  
MAGNESIUM-CHELATASE SUBUNIT CHLD PRECURSOR (MG-PROTOPORPHYRIN IX  
CHELATASE) (MG-CHELATASE SUBUNIT D).  
CHLD.  
GN Nicotiana tabacum (Common tobacco).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
NCBI\_TaxID=4097;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=98079235; Pubmed=9418040;  
RX Papenbrock J., Graefe S., Kruse E., Haenel F., Grimm B.;  
RA "Mg-chelatase of tobacco: identification of a Chl D cDNA sequence  
RT encoding a third subunit, analysis of the interaction of the three  
RT subunits with the yeast two-hybrid system, and reconstitution of the  
RT enzyme activity by co-expression of recombinant Chl D, Chl H and Chl  
RT I.";  
RL Plant J. 12:981-990(1997).  
RL  
CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES  
CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-  
CC PROTOPORPHYRIN IX.  
CC -1- PATHWAY: CHLOROPHYLL BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.  
CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 WFA DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: Y10022; CAA71128.1; -  
DR InterPro: IPR000523; Mg\_chelatase\_chII.  
DR InterPro: IPR002035; WFA.  
DR Pfam: PF01078; Mg\_chelatase; 1.  
DR SMART: SM00327; WFA; 1.  
DR PROSITE: PSS0234; WFA; 1.  
KW Photosynthesis; Chlorophyll biosynthesis; Chloroplast;  
KW Transit peptide.  
FT TRANSIT 1 62 CHLOROPLAST (POTENTIAL).  
FT CHAIN 63 758 MAGNESIUM-CHELATASE SUBUNIT CHLD.  
FT DOMAIN 556 752 WFA.  
FT DOMAIN 399 445 GLU/PRO-RICH.  
FT DOMAIN 409 416 POLY-PRO.  
FT DOMAIN 427 434 POLY-GLU.  
SQ SEQUENCE 758 AA; 83012 MW; ED531E85D5FAFEF8 CRC64;  
  
Query Match 6.2%; Score 87; DB 1; Length 758;  
Best Local Similarity 21.6%; Pred. No. 13;  
Matches 50; Conservative 3%; Mismatches 93; Indels 54; Gaps 9;

[illegible]

```

RESULT 14
STRP_STRQ      STANDARD;          PRT;   440 AA.
ID   STRP_STRQ      STANDARD;          PRT;   440 AA.
AC   P00779;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   STREPTOKINASE C PRECURSOR.
GN   SKC.
OS   Streptococcus equisimilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=119602;
RX   [1]
RN   SEQUENCE FROM N.A.
RC   STRAIN=H46A;
RX   MEDLINE=85232082; PubMed=2989113;
RA   Malke H., Roe B., Ferretti J.J.;
RT   "Nucleotide sequence of the streptokinase gene from Streptococcus
RT   equisimilis H46A.";
RL   Gene 34:357-362(1985).
[2]
RN   SEQUENCE OF 27-440.
RP   MEDLINE=83127125; PubMed=6760891;
RA   Jackson K.W., Tang J.;
RT   "Complete amino acid sequence of streptokinase and its homology with
RT   serine proteases.";
RL   Biochemistry 21:6620-6625(1982).
CC   -!- FUNCTION: THIS PROTEIN IS NOT A PROTEASE, BUT IT ACTIVATES
CC   PLASMINOGEN BY COMPLEXING WITH IT. AS POTENTIAL VIRULENCE FACTOR,
CC   IT IS THOUGHT TO PREVENT THE FORMATION OF EFFECTIVE FIBRIN
CC   BARRIERS AROUND THE SITE OF INFECTION, THEREBY CONTRIBUTING TO THE
CC   INVASIVENESS OF THE CELLS.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; K02986; AAA26974.1; -.
DR   EMBL; X72832; CAA51351.1; -.
DR   PIR; A00967; BZSO.
DR   PIR; A22801; A22801.
KW   Plasminogen activation; Signal; Virulence.
FT   SIGNAL          1 26
FT   CHAIN           27 440    STREPTOKINASE C.
FT   VARIANT         195 195    L -> D.
FT   VARIANT         207 207    D -> L.
FT   CONFLICT        298 300    EKY -> LEYK (IN REF. 2).
FT   CONFLICT        438 438    N -> D (IN REF. 2).
SQ   SEQUENCE      440 AA; 50140 MW; 8FC1F22648ACC77A CRC64;
Query Match          6.2%; Score 86.5; DB 1; Length 440;
Best Local Similarity 24.9%; Pred. No. 6.9;
Matches 56; Conservative 31; Mismatches 87; Indels 51; Gaps 10;

Qy 52 DDLMOVIDGASRILERSLSRSRVIANVYFPHPDTPGVTLTADPV-----VFQR 103
|| :||| || : : : || :
Db 121 DDYFEVIDFAS---DATIDRN-----GKYVFADKD-GSVTLPTQPVQEFLLSGHVRV 170
|| :||| || : : : || :

Qy 104 ELRQLYVGGGDCPDMVGAKAAVEVANPCSFYVFSDAKADYHKNNELLQLQLKQ- 162
: : : || : || : || : || : || : || : || : || : || : || : || : || :
Db 171 PYKEKPIQNA-----KSDVEYTVQGFPLNDDDFRPGDKTKLLKTLAIGDT 219
|| :||| || : || : || : || : || : || : || : || : || : || : || : || :
Qy 163 --SÖVFLVLTGCDGRTHPGYLAFAEIAS--TSSGQFOLDKQOVSEVLKWWESAIAQSK 218
|| :||| || : || : || : || : || : || : || : || : || : || : || : || :
Db 220 ITSQELLAQAOSILKNHPGTYTVERDSSIVTHNDIFRTLPLMDQEFYRVKNREQAYR 279
|| :||| || : || : || : || : || : || : || : || : || : || : || : || :

Qy 219 VHLISADHEE-----EGEHTWRIPFDPDS-LKEVTI 247

```



Search completed: March 7, 2002, 22:13:04  
Job time: 457 sec

...

GenCore version 4.5.  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 21:15:02 ; Search time 65.89 Seconds  
(without alignments)  
610.486 Million cell updates/sec

Title: US-09-665-728-1  
Perfect score: 1403  
Sequence: 1 MTPRAQLLLIATYVWAA.....IEVRDPLGMSQSPPLLMQD 275

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhmc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp.vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	628.5	44.8	5198	5 076518	O76518 caenorhabdi
2	282.5	20.1	536	4 Q9UMP9	Q9UMP9 homo sapien
3	282.5	20.1	852	4 Q9Y334	Q9Y334 homo sapien
4	282.5	20.1	873	11 Q9Z108	Q9Z108 mus musculu
5	282.5	20.1	891	11 Q9JHA8	Q9JHA8 mus musculu
6	116	8.3	353	11 Q9QXEL	Q9QXEL mus musculu
7	110	7.8	2229	5 Q19853	Q19853 caenorhabdi
8	109	7.8	2014	5 Q22774	Q22774 caenorhabdi
9	106	7.6	2470	13 Q90681	Q90681 gallus gall
10	105.5	7.5	1363	1 Q9HNT5	Q9HNT5 halobacteri
11	105	7.5	476	5 Q23410	Q23410 caenorhabdi
12	105	7.5	2214	5 Q20219	Q20219 caenorhabdi
13	101.5	7.2	1324	2 Q52048	Q52048 streptomyce
14	96.5	6.9	1292	10 Q9M1Q9	Q9M1Q9 arabidopsis
15	93.5	6.7	1576	2 Q05201	Q05201 rhodococcus
16	93.5	6.7	1576	2 Q69354	Q69354 rhodococcus
17	93	6.6	458	5 Q93377	Q93377 caenorhabdi
18	93	6.6	861	11 Q9EQK5	Q9EQK5 mus musculu
19	93	6.6	1362	1 Q9V294	Q9V294 pyrococcus

20	92.5	6.6	510	8 Q9G8W0	Q9G8W0 rhodomonas
21	92	6.6	584	2 Q9XCV7	Q9XCV7 streptomyce
22	92	6.6	611	5 Q9NLD1	Q9NLD1 caenorhabdi
23	92	6.6	788	13 Q07012	Q07012 xenopus lae
24	92	6.6	1648	11 Q9Q205	Q9Q205 mus musculu
25	92	6.6	1944	5 Q9N375	Q9N375 caenorhabdi
26	91.5	6.5	414	2 Q53284	Q53284 streptococc
27	91	6.5	1250	5 Q23409	Q23409 caenorhabdi
28	90.5	6.5	392	10 Q40742	Q40742 oryza sativ
29	89.5	6.4	772	13 Q9PU04	Q9PU04 ictalurus p
30	89	6.3	537	2 Q9AB06	Q9AB06 caulobacter
31	89	6.3	1716	5 Q9VZG5	Q9VZG5 drosophila
32	88.5	6.3	547	2 Q9KR88	Q9KR88 vibrio chol
33	88.5	6.3	869	10 Q9FW67	Q9FW67 oryza sativ
34	88	6.3	403	4 Q43460	Q43460 homo sapien
35	88	6.3	789	5 Q9BKQ4	Q9BKQ4 caenorhabdi
36	88	6.3	1229	10 Q49749	Q49749 arabidopsis
37	88	6.3	1229	10 Q9SY12	Q9SY12 arabidopsis
38	88	6.3	1229	10 Q9FWX8	Q9FWX8 arabidopsis
39	88	6.3	1278	10 Q9FWX7	Q9FWX7 arabidopsis
40	87.5	6.2	552	1 Q28762	Q28762 archaeoglob
41	87	6.2	353	2 P72409	P72409 streptomyce
42	87	6.2	417	11 Q9Z0R1	Q9Z0R1 cavia porce
43	87	6.2	791	6 Q9GK49	Q9GK49 bos taurus
44	87	6.2	3722	2 P94873	P94873 lysobacter
45	86.5	6.2	1230	2 Q59325	Q59325 clostridium

ALIGNMENTS

RESULT 1  
O76518 PRELIMINARY;  
ID O76518 Q10036; PRT: 5198 AA.  
AC O76518 Q10036;  
DT 01-NOV-1998 (TREMREL. 08, Created)  
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)  
DE HEMICENTIN PRECURSOR.  
GN HIM-4 OR F15G9.4  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Sulston J., Hedgecock E.M.;  
RT "Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and  
RT Germ-Line Chromosome Segregation in C. elegans.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Sulston J., Kershaw J.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM F15G9.4A AND ISOFORM  
CC F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
DOMAIN.  
DR EMBL; AF074901; AAC26792.1; -;  
DR EMBL; 247068; CAA87336.1; -;  
DR EMBL; 247070; CAA87336.1; JOINED.  
DR EMBL; 247070; CAA87345.1; -;  
DR EMBL; 247068; CAA87345.1; JOINED.  
DR EMBL; 247068; CAA87335.1; -;  
DR EMBL; 247070; CAA87335.1; JOINED.  
DR EMBL; 247070; CAA87344.1; -;  
DR EMBL; 247068; CAA87344.1; JOINED.  
DR HSSP; P35555; IEMN.  
DR WormPep; F15G9.4A; CE18595.  
DR WormPep; F15G9.4B; CE18596.  
DR InterPro; IPR000152; Asx\_hydroxyl.

DR InterPro: IPR000515; BPD\_transp.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF-Ca.  
DR InterPro: IPR003598; Ig\_C2.  
DR InterPro: IPR003600; Ig\_Like.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR002035; VNFA.  
DR Pfam: PF00047; Ig\_47.  
DR SMART: SM00179; EGF\_CA; 1.  
DR SMART: SM00001; EGF\_Like; 1.  
DR SMART: SM00408; IGC2; 44.  
DR SMART: SM00410; IGLike; 3.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
DR PROSITE: PS00402; BPD\_TRANS INN\_MEMBER; UNKNOWN\_1.  
DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
DR PROSITE: PS01187; EGF\_CA; 2.  
KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;  
KW Repeat; Signal.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 5198 HEMICENTIN.  
FT VARSPIC 1462 5198 MISSING (IN ISOFORM FI5G9\_4A).  
SQ SEQUENCE 5198 AA; 570816 MW; DA85111F2B58D37B CRC64;

Query Match 44.8%; Score 628.5; DB 5; Length 5198;  
Best Local Similarity 48.6%; Pred. No. 1.7e-43;  
Matches 126; Conservative 51; Mismatches 69; Indels 13; Gaps 3;

QY 10 LLLATYVVAAVTSDEPTKTLSPATGDTALAFVDFVTGSMWDDLMOVLDGASRLERSL 69  
DB 15 LLLAT---TCSSVNDK-----NDPTGKSSLAFAVDFITGSMFDDLVQVREGAKIFKTYM 66  
QY 70 SSRSRVIANVYALPHDPDIPGVTLTADPVFQRELRLQLYVGGDGPMSVGAIAAAYE 129  
DB 67 AQRELIYNYINVPHDPVLGLINTDSTYFNRLSKVYVHGGDCEPKTITGLIKALQ 126  
QY 130 VANPGSFYVDSARAKYHKKNELQLQLKQSVFVLTGDCGDRTHPGYLAFEEIAS 189  
DB 127 ISLPSFSYFVTDARKYHLEDEVLNTIQEQSSVFMVMTGCGNRTHPGERTVEKIAA 186  
QY 190 TSSGVFOLDKQVSEVLKWSAIOAKSVHLLSADHEEGHTWRIPDPSLKVEVTISL 249  
DB 187 ASFGQVHLEKSDVSTVLEYVHVAVKQKVHLMYERARGVTVSRNIPVDKHLSELTISL 246  
QY 250 SG-----PGPEIEVRDPLG 263  
DB 247 SGDKDDSDNLDIVLRDPEG 265

RESULT 2  
ID Q9UMP9 PRELIMINARY; PRT; 536 AA.  
AC Q9UMP9;  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)  
DE G7C PROTEIN (FRAGMENT).  
GN G7C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Aguado B.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Albertella M.R.; Campbell R.D.;  
RT "Characterisation of the novel gene G7c located in the class III  
region of the human Major Histocompatibility Complex.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96411681; PubMed=8812450;  
RA Albertella M.A.; Jones H.; Thomson W.; Olavesen M.G.; Campbell R.D.;  
RT "Localization of eight additional genes in the human major  
histocompatibility complex, including the gene encoding the casein  
kinase II beta subunit (CSNK2B).";  
RL Genomics 36:240-251(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Albertella M.R.;  
RL Thesis (1997); University of Oxford, Department of Biochemistry,  
MRC Immunochemistry Unit.  
DR EMBL: AJ245418; CAB52192.1; -.  
FT NON\_TER 1  
FT NON\_TER 536  
SQ SEQUENCE 536 AA; 57837 MW; B32D8DD8E24BD06D CRC64;

Query Match 20.1%; Score 282.5; DB 4; Length 536;  
Best Local Similarity 29.2%; Pred. No. 1.3e-15;  
Matches 79; Conservative 50; Mismatches 93; Indels 49; Gaps 7;

QY 25 DEPTKTLSPATGDTALAFVDFVTGSMWDDLMOVLDGASRLERSLSRSR-----VIANYA 80  
DB 3 DRDFSRLLDITPASSLSFLVLDITGSMGEI-----NAKIQARHLVQRRGSPMPVHYV 57  
QY 81 LVPHDPDIPGVTLTADPVFQRELRLQLYVGGDGPMSVGAIAAAYEVNPGSTIYVF 140  
DB 58 LVPHDPGPGPVFTTSDPDSFWQLNEITHALGGDEPCMLSAQLALLHTPLSLDIFVF 117  
QY 141 SDARADYHKKNELQLQLKQSVFVLTGDCG---DRTHPGYLA-----FEETASNS 191  
DB 118 TDASPKDAFLTNOVSLTQERRCRVTLVTESTRVQGRARRRILSPLRFPYKAVLAS 177  
QY 192 SGQVQLDKQVSEVLKWSAIOAKSVHLLSADHEEGHTWRIPDPS-----241  
DB 178 GGEVIFTKQHIRDVAIVGESMAA---LVT-----LPLDPPVVVPGQPLVF 221  
QY 242 -----LKEVTISLSGPGPEIEVRDPLGMSQG 267  
DB 222 SYDGLLQKITVRIHGDISSFWINKPAGVSQG 252

RESULT 3  
ID Q9Y334 PRELIMINARY; PRT; 852 AA.  
AC Q9Y334;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)  
DE NG37.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rowen L.; Qin S.; Madan A.; Dickhoff R.; Dors M.; Madan A.; Hicks P.;  
RA Loretz C.; Ratcliffe A.; Abbasi N.; Shaffer T.; Hood L.;  
RT "Sequence of the human major histocompatibility complex class III  
region.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF134726; AAD21820.1; -.  
SQ SEQUENCE 852 AA; 92538 MW; 025430B3F912A941 CRC64;

Query Match 20.1%; Score 282.5; DB 4; Length 852;  
Best Local Similarity 29.2%; Pred. No. 2.5e-15;  
Matches 79; Conservative 50; Mismatches 93; Indels 49; Gaps 7;

QY 25 DEPTKTLSPATGDTALAFVDFVTGSMWDDLMOVLDGASRLERSLSRSR-----VIANYA 80  
DB 3 DRDFSRLLDITPASSLSFLVLDITGSMGEI-----NAKIQARHLVQRRGSPMPVHYV 57

```

301 DROFSRLDITPASSLSFVLDITGSGEEI-----NAAKIQARHLVEQRGSGMEPVHV 355
Qy 81 LVPHDPDIPGVILTADPVVVFQRELRLQLYVQGGDCPEMSVGAIKAAVEVANPGSFIYV 140
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 LVPHDPDIPGVFVITSDPDSFWOQLNEIHAIHGGDEPEMCLSAIQALLHTPTPLSDIFV 415
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 141 SDARAKYHKHKNELLQLLQLKQSQVVFVLTGDCG-----DRTHPGYLA-----FEETAS 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 TDASPKDAFLTNQVESLTQERCCRVFLVITEDTSRVQGRARREILSPLEFEPYKAVALAS 475
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 192 SGQVFQLDKQGVSEVLKWKVESATQASKVHLLLSADHEEGEHTWRIPDPFS----- 241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 GGEVIFTKQHIIRDVAIVGESMAA-----LVT-----LPLDPPVVVVGQPLVF 519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 242 -----LKEVTISLSGGPGPEIEVRDPLGMSOG 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 SVDCGLLQKITVRINGDISSEFWIKNPAGVSGO 550
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
Q9Z108 PRELIMINARY; PRT; 873 AA.
AC Q9Z108;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 94.9 KDA PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J29;
RA Rowen L., Madan A., Qin S., Shaffer T., Ratcliffe A., Abbasi N.,
RA Dickschhoff R., James R., Loretz C., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
RT region.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF109505; AAC84152.1; -.
DR InterPro: IPR002035; VFWA.
DR SMART: SM00327; VWA; 1.
DR HYPOTHETICAL protein.
SQ SEQUENCE 873 AA; 94894 MW; 5761E3620F8D831C CRC64;

```

Query Match	20.1%;	Score	282.5;	DB 11;	Length	873;			
Best Local Similarity	30.3%;	Pred. No.	2.6e-15;						
Matches	81;	Conservative	45;	Mismatches	100;	Indels	41;	Gaps	
Qy	25	DEPTKILSPATGATIAFAFVDTVGTSGMWDDLQMQVIDGASRILERSLSRSRVNTIANVALVPF	84						
Db	310	DKAFSRLLDITPASSLSFVLDTTTSGMEINAAKIQARRIVEQRQSGPMEPVF-YILVPF	368						
Qy	85	HDPDIGPVTITADPVVFORLRLQYVQGGSDCPMSVSGAIKAAEVANVPGSTIYVFS	144						
Db	369	HDGFGVFTTSDPDSFQWKLNIHALGGDEPMCLSALEALLHTPPLSDIFVTDAS	428						
Qy	145	AKDYHKKNELLQLLQKQSOVFLVLTGD---CDRTHPGYLA-----FEEIASSTSSGOV	195						
Db	429	PKDALLTNRVESLTERRRCHVTLFTVDEPSRTGRRREALSPLRPEEYETARASGGEV	488						
Qy	196	FQLDKQOVSEVLKWEISAIQAKVHLLSADHEEGEHTWRIIPDPFS-----	241						
Db	489	IFTKDQYIQD-----VAAIVGESMAGLVT-----LPLDPVFTPGPCVFSVDS	532						
Qy	242	LKAEVTISLSPGPEIEVRDPLGMSOG	267						
Db	533	LLMQVTVRMHGDISSFWIKSPAGVSG	559						

RESULT  
Q9JHA8  
5

[illegible]

GN SHD-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Yoshida M., Kuwahara K., Sakaguchi N.;
RT "Molecular cloning of a novel gene, SHD-1, belonging to cell cycle-
RT related SAC3 family.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131957; CAB65242.1; -
SQ SEQUENCE 353 AA; 38203 MW; D486B7B301493447 CRC64;

Query Match      8.3%; Score 116; DB 11; Length 353;
Best Local Similarity 24.9%; Pred. No. 0.089;
Matches 70; Conservative 31; Mismatches 98; Indels 82; Gaps 13;

Qy 3 PRAQLL---PLLATYTVVAAVTSDEPTKLTSLPATGDATLA---FVFDVTGSMWDDL- 54
Db 64 PPSLLRPPVLLATVRYLAGEVA-----GRGDVSCAEVASFVADRLRAVRLDLS 113

Qy 55 MQVID--GASRILERSRSRVIANYPHDPDIPGVTL--TADPVVFQRE----- 104
Db 114 LOGVDADAATVLEAALATLLAVAR-----VRPEETRGAAADPVLLQTQVQEGFG 163

Qy 105 -LRQLYVOGGDCP-----EMSVGATKAAVEVANPGSFYFVFSARAK 146
Db 164 SURRCYARGKGPYRQAFAOGLFLLYNLGSLAPAGGTAALRACPLQAALAVDAARE 223

Qy 147 DYHKK-NELLQLQLKQSVFVLTGDCGDRTHPGYLAFEITA-----STSSGOVFOLD 199
Db 224 DNHARLRLRLTLPYQS-----CAVQEHGYARKALARLSRALSTPKGTQLPLD 274

Qy 200 KQVSEVLKVESATQASKVHLLSAD-----HEEG 230
Db 275 FTEHFLALDGLQEARLDLCOAHGLTLDKRVVFLRQYSEEG 315

RESULT 7
Q19853 ID Q19853 PRELIMINARY; PRT; 2229 AA.
AC Q19853;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SIMILAR TO EGF-LIKE REPEAT. NCBI GI: 1125759.
GN F28B4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardiner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders B., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RA Waterston R.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42834; AAA83584.1; -
DR InterPro; IPR001442; C4.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001304; lectin_c.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00111; C4; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS0041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS02034; VWFA; 1.
DR EGF-like domain; Glycoprotein.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2229 AA; 244874 MW; 873C82B6A97AEA25 CRC64;

Query Match      7.8%; Score 110; DB 5; Length 2229;
Best Local Similarity 20.1%; Pred. No. 4.4; 119; Indels 40; Gaps 9;
Matches 52; Conservative 48; Mismatches 119; Indels 40; Gaps 9;

Qy 39 TLAFVFDVTGSMWDDLMOQVIDGASRILERSRSRVIANYPHDPDIPGVTL----- 94
Db 1551 TLLFVVEITRQSDTVDNQLIANKNIVTSA-----TNFA--PFWSEYFGLVTFDITG 1600

Qy 95 -----TADPVVFQRELRLQLYVOGGDCPEMSVGAIKAAVE-----VANPGSFYFVS 141
Db 1601 RTFEKYNVTITDALITDLTAQSTAISTDGCASMPYLGVLHLLHEDHNVISIPNSEIFLVT 1660

Qy 142 DARAKDYHKKNELLQLKQSVFVLTGDCGDRTHPGYLAFE-----IASTSSGOVFQL 198
Db 1661 AAGPSDLNKYGEAMNSLENTEAHLHYIVSKSANCPTEFGVNVNQDMTWLGCGSSGNILFT 1720

Qy 199 DKQVSEVL-KWVESATQASKVHLLS-----ADHE-EEGEHTWRIPDPSPKKEVTISLG 251
Db 1721 DSSNIVSLMNSYLPISLYGAS---ILQDPTGPANYSCTDGLSPWFVPVDSNTTFFIVTTSS 1777

Qy 252 PGPEIEVRDPLGMSOGSPP 270
Db 1778 EFGSLSVKDLPLQAHNVAP 1796

RESULT 8
Q22774 ID Q22774 PRELIMINARY; PRT; 2014 AA.
AC Q22774; O01703; O01984;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEICAL 221.3 KDA PROTEIN T25C12.3 IN CHROMOSOME X.
GN T25C12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Kershaw J.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WEAK, TO C.ELEGANS ZK1193.2 AND TO THE C-TERMINAL OF
CC COLLAGENS.
DR EMBL; Z66566; CAA91487.1; -
DR EMBL; Z78543; CAA91487.1; JOINED.
DR EMBL; Z93779; CAA91487.1; JOINED.
DR EMBL; Z93779; CAB07849.1; -
DR EMBL; Z78543; CAB07849.1; JOINED.
DR EMBL; Z66566; CAB07849.1; JOINED.
```



```

DR SMART: SM00059; FN2: 1.
DR PROSITE: PS00023; FIBRONECTIN_2; UNKNOWN_1.
DR PROSITE: PS00149; SULFATASE-2; UNKNOWN_1.
DR SEQUENCE 2470 AA; 275645 MW; DB33824479CDA13 CRC64;
      8;

Query Match      7.6%; Score 106; DB 13; Length 2470;
Best Local Similarity 23.4%; Pred. No. 11;
Matches 43; Conservative 30; Mismatches 59; Indels 52; Gaps

QY 15 YTVVAAAVTSD--EPTKTLSPATGATLAFVFDVTGSMDDLMQVIDGASRLERSLSR 72
   | | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 1478 FTAACPLKSNVQDCRVMPATG-----HLFDLTSLKRESGYITDShNRKIELNVCAE 1532

QY 73 SR-VIANALVPFHPDIPGVTLTATA---DPVVQRELRLQLYVGGGDCPEMSVGAIAKAA 127
   : : | | | | | : | | | | : : : : : | | | | : : : : :
Db 1533 AKSCANGAAVCTID---GPKTLNAGKSLTLYTDVQKLVIEDGQPCP----- 1579

QY 128 VEANPGSFYVFSARAKDYHKKNELLQLQLKQSQVFFLTGDCGDRTHPGYLAPEEI 187
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1580 -----TDLATK--HK-----SYFSFCKSDAGDSDQPVFLSFDEQ 1612

QY 188 ASTS 191
   ||
Db 1613 TCTS 1616

RESULT 10
Q9HNT5 PRELIMINARY; PRT; 1363 AA.
ID Q9HNT5
AC Q9HNT5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VNG1953C.
DE VNG1953C.
GN Halobacterium sp. (strain NRC-1).
OS Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; Pubmed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Erhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005091; AAG20135.1; -.
KW Complete proteome.
SQ SEQUENCE 1363 AA; 138692 MW; A34770E00AA5A58 CRC64;

Query Match      7.5%; Score 105.5; DB 1; Length 1363;
Best Local Similarity 22.2%; Pred. No. 5.1;
Matches 58; Conservative 33; Mismatches 91; Indels 79; Gaps 10;

QY 13 ATYTVAAAVTSDP-----FKTLSPATGATLAFVFDVTGSMDDLMQVIDGASRI 64
   | | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 298 AASMAITADATDTPLVVPVTPSTRVNSTDALDVAYRN-----QPVSGGSAD 347

QY 65 LERSLS-----SRSRVIAN-----YAL-VPEHPDIPGV 92
   : | | | | : | | | | : | | | | : | | | | : | | | |
Db 348 VSMTLAANGTENATFDISDPLVANATQRLTDLAENAPLADGVYTLSTVTDGSGSPV 407

QY 93 TLTADPVVQRELRLQLYVGGGDCPEMSVGAIAVEVANPGSFI---YVFSARAKDYH 149
   | | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 408 TRTTPDVVIDD-----EPDPVSAVLSNASTDV--SPRDTVAVNNTYDDA----- 450

```

```
QY 150 KKNELLQLLQKQSOVVFVLTGCGDRTHPGYLAFPEIASTSSGOVQFQDLKQKVSEVLKW 209
Db 451 -----TNATSATVHFVAADAGNFT-AGDLANASASVSKDIDMAPGCKORVDVNLRW 501
QY 210 VESAIQASKVHLLSADHEEG 230
Db 502 VREVTDSNVTYVVTATDERG 522

RESULT 11
Q23410
ID Q23410 PRELIMINARY; PRT; 476 AA.
AC Q23410;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE SIMILARITY OVER A SHORT REGION TO TENASCIN PRECURSORS.
GN ZK1193.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R.; Alnscough R.; Anderson K.; Baynes C.; Berks M.; Coulson A.;
RA Bonfield J.; Burton J.; Connell M.; Copsey T.; Cooper J.; Coulson A.;
RA Craxton M.; Dear S.; Du Z.; Durbin R.; Favello A.; Fulton L.;
RA Gardner A.; Green P.; Hawkins J.; Hillier L.; Jier M.; Johnston L.;
RA Jones M.; Kershaw J.; Kirsten J.; Laister N.; Latreille P.;
RA Lightning J.; Lloyd C.; McMurray A.; Mortimore B.; O'Callaghan M.;
RA Parsons J.; Percy C.; Rifken L.; Roopra A.; Saunders D.; Showkeen R.;
RA Smaldon N.; Smith A.; Sonhammer E.; Staden R.; Sulston J.;
RA Therry-Mieg J.; Thomas K.; Vaudin M.; Vaughan K.; Waterston R.;
RA Watson A.; Wellstock L.; Wilkinson-Sproat J.; Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Geisel C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41553; AAA83292.1; -.
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 476 AA; 53310 MW; F39C14163CC074DA CRC64;

Query Match 7.5%; Score 105; DB 5; Length 476;
Best Local Similarity 22.3%; Pred. No. 1.2;
Matches 53; Conservative 49; Mismatches 100; Indels 36; Gaps 10;

QY 38 ATAFVFDVTSMMDDLMQVIDGASRLIE-----RSLSSRSRVIANVLPVPHD 86
Db 82 SAIAFVIRSSSKAQLNEAEATNIVEYELHYFFQKFIILT---VVSNNAITFSHE 138
QY 87 PDIGFVLITADPVVQRELRLQYV-OGGDCPCPEMSVGAIAKAAVEA----NPGSFIYVFS 141
Db 139 YDIGE-----DFANSIRSLVAPPTETECDDALLAGISKTLENSAFKYPNSPVFVS 190
QY 142 DARAK-DYHKKNELQLQLKQSOVVFVLT-----GDCG-DRTHPGYLAFPEIASTSSGOV 195
Db 191 DCTANDFTTAGFQMEQIVNTRAILFMITESAGSGCNVDVSTNIFESRLSLQSLRGLL 250
QY 196 FQDKQKVSEVLKWKVESAIQASKVHLLSADHEEGEHTWRIPF--DPSLKEVTISLSG 251
```

```
Db 251 IOTSLMQLSDSFSAQDLWQYDT-ILTNLDECKRPMFQPFVQSDIFLTLRASG 307

RESULT 12
Q20219
ID Q20219 PRELIMINARY; PRT; 2214 AA.
AC Q20219;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA YK131E8.3.
GN F40F4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R.; Alnscough R.; Anderson K.; Baynes C.; Berks M.; Coulson A.;
RA Bonfield J.; Burton J.; Connell M.; Copsey T.; Cooper J.; Coulson A.;
RA Craxton M.; Dear S.; Du Z.; Durbin R.; Favello A.; Fulton L.;
RA Gardner A.; Green P.; Hawkins J.; Hillier L.; Jier M.; Johnston L.;
RA Jones M.; Kershaw J.; Kirsten J.; Laister N.; Latreille P.;
RA Lightning J.; Lloyd C.; McMurray A.; Mortimore B.; O'Callaghan M.;
RA Parsons J.; Percy C.; Rifken L.; Roopra A.; Saunders D.; Showkeen R.;
RA Smaldon N.; Smith A.; Sonhammer E.; Staden R.; Sulston J.;
RA Therry-Mieg J.; Thomas K.; Vaudin M.; Vaughan K.; Waterston R.;
RA Watson A.; Wellstock L.; Wilkinson-Sproat J.; Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Wilson R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40420; AAA81430.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00032; vwa; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00041; C-TYPE-LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS02334; VWFA; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2214 AA; 243435 MW; B0B387B67013A401 CRC64;

Query Match 7.5%; Score 105; DB 5; Length 2214;
Best Local Similarity 19.4%; Pred. No. 12;
Matches 51; Conservative 46; Mismatches 120; Indels 46; Gaps 8;

QY 39 TLAFFVDVTSMMDDLMQVIDGASRLIE-----VANPGSFIYVFSDAKADY 148
Db 1548 SLLMVVETTRQNSDTVNQLIANKNIVSATVNNMPLMFTNFGLVTF---DTTGRTFEKF 1604
QY 95 --TADPVVFORELRLQYVGGGDCPEMSVGAIAKAAVE---VANPGSFIYVFSDAKADY 148
Db 1605 YTSIDDLITLTITQSNIAISTDGVCSMPYLGVLHLLHDDVDVIAMPNSEIFLVTPAGPSDL 1664
QY 149 HKKNELQLQLKQSOVVFVLT--TGDCG-----DRTHPGYLAFPEIASTSSGOV 197
```

Db 1665 GNYVETMEVLNFOAHLHYVSKTANCATFDGVNNVDMTWLGY-----GSSGNILF 1716

QY 198 LKQOQSEVLKQVESIAQSKV-----HLLSADHEEGEHTWRIPFDPSLKEVTIS 248

Db 1717 TDPANIVNLFNFYLPYGLYASVLQDPTGINTVTC-----DGLPWFVFPVDINTTIYVT 1771

QY 249 LSGPGIEVDRPLGMSQSPPL 271

Db 1772 TSAEFGSLVKOPLGAHSAATPV 1794

RESULT 13

052048 PRELIMINARY; PRT; 1324 AA.

AC 052048; 01-JUN-1998 (Tremblrel. 06, Created)

DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE PEPTIDE SYNTHETASE MODULE (FRAGMENT).

GN CPS-1.

OS Streptomyces fradiae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OX NCBI\_TaxID=1906;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A54145;

RA Hosted T.J., Baltz R.H.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF016696; AAC01566.1; -.

DR HSP; P14687; 1AMU.

DR InterPro: IPR000873; AMP-bind.

DR InterPro: IPR001242; DUF4.

DR InterPro: IPR003880; Phosphopant\_attach.

DR Pfam: PF00501; AMP-binding; 1.

DR Pfam: PF00668; Condensation; 2.

DR Pfam: PF00550; pp-binding; 1.

DR PRINTS; PR00154; AMPBINDING.

DR PROSITE; PS00075; ACP\_DOMAIN; 1.

DR PROSITE; PS00455; AMP\_BINDING; UNKNOWN\_1.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.

FT NON\_TER 1

FT 1324 1324

FT SEQUENCE 1324 AA; 140539 MW; 24285EDD3690964F CRC64;

Query Match 7.2%; Score 101.5; DB 2; Length 1324;

Best Local Similarity 20.6%; Pred. No. 11;

Matches 67; Conservative 38; Mismatches 123; Indels 97; Gaps 161;

QY 10 LLLATYVVAAVTSDEPTKTLSPAT-----GDATLAFVFDVTGSMWDDLMOV 57

Db 102 LLERALTAVSAPATPTAALNATPATARAAPRAAGCAPATVDAFEARVATPEAPAV 161

QY 58 IDG-----ASRLERSLSRSRV---IANYALVPF-----HDP 87

Db 162 LAGGEELTYAELDARANRLARLLERGVGPESRYALTVSRNWLPAVLGILKAGGCYVP 221

QY 88 -----DTGPVTLTADPVVFORELQRLVQGGDCPEMSVGAIKAAVEVAN 132

Db 222 VGATLPRERAARILRETAPVCLLTDPAEAAARTRT-APTGDGRDENAPGVE---RVVL 277

QY 133 PGSTIYVFSARAKDYHKKNELLOLLOLKQSOVVFVLTGCGDRTHPGYLAFAEIASTSS 192

Db 278 TGALLAAPPAPPPDAERAGPLL-----PCHLAILLHTSGSS 314

QY 193 GQV--FQLDKQVSEVLKQVESIAQSKVHLLSADHEEGEHTWRIPFDPSLKEVTISLS 250

Db 315 GRPKGVAVEHAQVTLASWAGTGVGADRLHRTVASTSES-----FDVSVFTLLVPLL 366

QY 251 GPGPEIEVRDPLGMSQ---GSPPLL 272

Db 367 TGGRIEIVENTLAVADRTGGEPSSL 391

RESULT 14

Q9MIQ9 PRELIMINARY; PRT; 1292 AA.

AC Q9MIQ9; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE P-GLYCOPROTEIN-LIKE PROTEIN.

GN T17J13.110.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,

RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC

CC TRANSPORTERS).

DR EMBL; ALI38651; CAB71875.1; -.

DR InterPro: IPR003593; AAA.

DR InterPro: IPR001140; ABC\_transporter\_tm.

DR InterPro: IPR003439; ABC\_transportr.

DR InterPro: IPR001687; ATP\_GTP\_A.

DR InterPro: IPR001525; C5\_DNA\_meth.

DR Pfam: PF00664; ABC\_membrane; 2.

DR Pfam: PF00005; ABC\_tran; 2.

DR SMART; SM00382; AAA; 2.

DR PROSITE; PS00211; ABC\_TRANSPORTER; 2.

DR PROSITE; PS00095; C5\_MTASE\_2; UNKNOWN\_1.

KW ATP-binding; Transport.

SQ SEQUENCE 1292 AA; 139753 MW; 9C95DF3AC84E7432 CRC64;

Query Match 6.9%; Score 96.5; DB 10; Length 1292;

Best Local Similarity 29.2%; Pred. No. 27;

Matches 42; Conservative 15; Mismatches 50; Indels 37; Gaps 7;

QY 24 SDEPTKTLSPATGDATLAFVFP-----DVTGSMWDDLMOVIDCASRI--LERSLSRSR 74

Db 1034 SDESGRVLNWKGDIELRHISFKYPSRPDV--QIFDLCLISIRAGKTIALVGGSGKST 1091

QY 75 VIANYALVPFHDPPDIPVTL-----TADPVVFORELQRLVQV-G 113

Db 1092 VIA--LLQRYDPSGGITLDGVEIKTLQKLWLRQQTGLVSOEVLNENIRANIAVGK 1149

QY 114 GDCPEMSVGAIKAAVEVANPGSFI 137

Db 1150 GDATETE---IVSAAELSNAGFI 1170

RESULT 15

005201 PRELIMINARY; PRT; 1576 AA.

AC 005201; 01-JUL-1997 (Tremblrel. 04, Created)

DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE RECEPTOR-LIKE HISTIDINE KINASE BPDS.

GN BPDS.

OS Rhodococcus sp.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.

OX NCBI\_TaxID=1831;

RN [1]

Search completed: March 7, 2002, 22:07:47  
Job time: 3165 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:05:20 ; Search time 48.12 seconds  
(without alignments)  
103.136 Million cell updates/sec

Title: US-09-665-728-1\_COPY\_56\_122

Perfect score: 342

Sequence: 1 QVTDGASRLERSLSRSRV.....RELRLYVGSGDCPEMSVG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SID52/gcgcdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID52/gcgcdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID52/gcgcdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID52/gcgcdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID52/gcgcdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID52/gcgcdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID52/gcgcdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID52/gcgcdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID52/gcgcdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID52/gcgcdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID52/gcgcdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID52/gcgcdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID52/gcgcdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID52/gcgcdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID52/gcgcdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID52/gcgcdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID52/gcgcdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID52/gcgcdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID52/gcgcdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID52/gcgcdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID52/gcgcdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID52/gcgcdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	100.0	275	22	AA831147
2	175	51.2	3117	21	AA53667
3	64	18.7	190	21	AA52665
4	64	18.7	229	21	AA52664
5	63.5	18.6	201	22	AA41239
6	63	18.4	215	21	AA41586
7	63	18.4	215	21	AA19587
8	63	18.4	223	22	AAE00828
9	63	18.4	225	19	AAW29715
10	63	18.4	225	19	AAW29716
11	63	18.4	225	19	AAW56141
					Rat secreted facto
					Sequence g1/32818
					Arabidopsis thalia
					Arabidopsis thalia
					Human polypeptide
					Human interleukin-
					Mouse interleukin-
					Human cardiographi
					Human neurotrophic
					Mouse neurotrophic
					Amino acid sequenc

12	63	18.4	225	19	AAW56142	Amino acid sequenc
13	63	18.4	225	20	AAW94466	Human cardiographi
14	63	18.4	225	21	AA57813	Human NNT-1 protei
15	63	18.4	225	21	AA57814	Murine NNT-1 prote
16	63	18.4	225	22	AA63543	Amino acid sequenc
17	63	18.4	253	22	AA635831	Human protein sequ
18	61	17.8	746	20	AA529512	Human lung tumour-
19	61	17.8	746	21	AA44461	Human lung tumour-
20	60	17.5	125	21	AA626109	Zea mays protein f
21	60	17.5	143	21	AA635542	Arabidopsis thalia
22	60	17.5	145	21	AA640813	Zea mays protein f
23	60	17.5	165	21	AA640812	Zea mays protein f
24	60	17.5	173	21	AA635541	Arabidopsis thalia
25	60	17.5	212	21	AA614557	Arabidopsis thalia
26	60	17.5	214	21	AA640739	Zea mays protein f
27	60	17.5	216	21	AA634742	Arabidopsis thalia
28	60	17.5	244	21	AA634741	Arabidopsis thalia
29	60	17.5	252	21	AA614556	Arabidopsis thalia
30	60	17.5	257	21	AA640738	Zea mays protein f
31	60	17.5	261	22	AAU00701	Mouse FCTR4 protei
32	60	17.5	274	20	AA502097	A multifunctional
33	60	17.5	274	20	AA502098	A multifunctional
34	60	17.5	370	21	AA648663	Mouse growth facto
35	60	17.5	370	22	AAU00700	Mouse FCTR3 protei
36	60	17.5	370	22	AA60895	Human VEGF-G prote
37	59.5	17.4	442	22	AA64895	Human protein sequ
38	59	17.3	113	21	AA625892	Zea mays protein f
39	58	17.0	93	21	AA641201	Zea mays protein f
40	58	17.0	248	21	AA635960	Zea mays protein f
41	57.5	16.8	362	10	AA694763	Formate dehydrodge
42	57	16.7	417	21	AA659010	Breast and ovarian
43	57	16.7	501	22	AA639590	Human polypeptide
44	57	16.7	516	22	AA641376	Human polypeptide
45	57	16.7	661	21	AA544927	Zea mays MAP kinas

ALIGNMENTS

RESULT 1						
AA83147	ID	AA883147	standard; protein; 275 AA.			
XX	AC	AA83147;				
XX	DT	29-JUN-2001	(first entry)			
XX	DE	Rat secreted factor encoded by clone P00210D09.				
XX	DE	Rat; secreted factor; P00210D09; cardiac; nephrotropic;				
XX	KW	antiinflammatory; gene therapy; cardiac disease; renal disease;				
XX	KW	inflammatory disease.				
XX	OS	Rattus sp.				
XX	PH	Key	Location/Qualifiers			
FT	Peptide	1..21				
FT	Protein	/label= Signal_peptide				
FT	Domain	22..275				
FT	Domain	/note= "Rat secreted factor"				
FT	Domain	35..55				
FT	Domain	/label= Transmembrane_domain				
FT	Domain	123..143				
FT	Domain	/label= Transmembrane_domain				
XX	WO200123419-A2.					
XX	05-APR-2001.					
XX	27-SEP-2000; 2000WO-US26582.					
XX	27-SEP-1999; 99US-0156277.					

```
PA (SCIO-) SCIOS INC.
XX
PI Stanton LW, Kapoun AM;
XX
DR WPI: 2001-328177/34.
XX N-PSDB: AAF82464.
XX
PT Novel secreted factor encoded by clone P00210D09 useful for diagnosing,
PT treating and/or preventing various cardiac, renal and inflammatory
PT diseases.
XX
PS Claim 9: Fig 1: 69pp; English.
XX
CC The present sequence is a novel secreted factor encoded by rat cDNA
CC clone P00210D09. The invention relates to a polypeptide comprising a
CC sequence of at least 80% identity to residues 22-122 of the present
CC sequence, or a sequence encoded by a nucleic acid hybridising under
CC stringent conditions to the complement of the coding region comprising
CC 1031 nucleotides, and having at least one biological activity of the
CC polypeptide encoded by clone P00210D09. The polypeptides and
CC polynucleotides of the invention are useful for the treatment of
CC cardiac, renal and inflammatory diseases. The polynucleotides are
CC useful in antisense mediated gene inhibition and in gene therapy.
CC The polypeptides are useful in assays for identifying lead compounds
CC that may be used as therapeutic agents in the treatment of cardiac,
CC kidney or inflammatory diseases.
XX
SQ Sequence 275 AA:

Query Match 100.0%; Score 342; DB 22; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.5e-38;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVIDGASRLERSLSRSRVIANVALVPFHPDIPGVTLTADPVVFORLRLYVQGGSD 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 qvidgasrlerslsrsrvianvalvpfhdipdipgtltadpvvfqreirqlvygggd 115
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 CPMSV 67
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 116 cpmsv 122

RESULT 2
AA53667
ID AAY53667 standard; Protein: 3117 AA.
XX
AC AAY53667;
XX
DT 22-FEB-2000 (first entry)
XX
DE Sequence gi/3328186 from an alignment with protein 608.
XX
KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KW bone development; gi/3328186.
XX
OS Unidentified.
XX
PN WO9960164-A1.
XX
PD 25-NOV-1999.
XX
PF 14-MAY-1999; 99WO-US11066.
XX
PR 15-MAY-1998; 98US-0085673.
XX
PA (QUARK-) QUARK BIOTECH INC.
XX
PI Einat P, Mor O, Skalliter R, Feinstein E, Faerman A;
XX WPI: 2000-053304/04.
XX
PT Identification of stress induced genes for determining risk and

preventing, treating or controlling osteoporosis.
XX
Claim 32: Fig 6A-R; 308pp; English.
XX
CC The present sequence is obtained from a clustral X alignment with
CC protein 608. Protein 608 was identified using the method of the invention
CC after subjecting rat osteoblasts to mechanical stress. Expression of the
CC 608 gene was found to be upregulated by about 3-fold in cells subjected
CC to mechanical strain. The specification describes a method for the
CC identification of genes responsive to a specific mechanical stress. The
CC method comprises applying the mechanical stress to an organism (tissue
CC or cells comprising bone cells), isolating the specific cellular
CC fractions and extracting mRNA from them, and differentially analysing the
CC mRNA in comparison with control samples. The method is used to identify
CC genes whose expression is responsive to a specific stress. The identified
CC genes are employed in determining risk associated with a physiological or
CC disease state. The risk determination methods are used for testing a
CC medicament for gene therapy. These medicaments, or genes identified by
CC the method of the invention, are used for treating, preventing or
CC controlling a physiological or disease state (especially osteoporosis or
CC bone density or other factors causing or contributing to osteoporosis or
CC its symptoms or other conditions involved in mechanical stress or its
CC lack. The methods can also be used for advancing research or studies in
CC bone development.
XX
SQ Sequence 3117 AA:

Query Match 51.2%; Score 175; DB 21; Length 3117;
Best Local Similarity 45.5%; Pred. No. 2.2e-14;
Matches 30; Conservative 16; Mismatches 20; Indels 0; Gaps 0;

QY 1 QVIDGASRLERSLSRSRVIANVALVPFHPDIPGVTLTADPVVFORLRLYVQGGSD 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 53 qvregaakiftvmaqrkellynylmvpfhdipylgeinttdstvfmrqlskvyvhgggd 112
QY 61 CPMSV 66
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 113 cpektl 118

RESULT 3
AAG52665
ID AAG52665 standard; Protein: 190 AA.
XX
AC AAG52665;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66973.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
```



```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      18.7%; Score 64; DB 21; Length 190;
Best Local Similarity 42.9%; Pred. No. 1;
Matches 21; Conservative 5; Mismatches 17; Indels 6; Gaps 2;

QY      6 ASRILERSLSRSRYANVALYFPHDPDGPVTLTADPVPVQRELRLQY 54
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      15 aprlnerllisslr--rsvaahpwhdlelgp----gapqifnvvdrlly 57

RESULT      4
AAG52664
ID AAG52664 standard; Protein; 229 AA.
XX
AC AAG52664;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66972.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
```



```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151920.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.7%; Score 64; DB 21; Length 229;
Best Local Similarity 42.9%; Pred. No. 1.3;
Matches 21; Conservative 5; Mismatches 17; Indels 6; Gaps 2;

QY 6 ASRLERSLSRSRVYANYALVPHDPDIPGVTLTADPVVFORELROLY 54
   ||| ||| ||| : ||| ||| ||| ||| ||| |||
Db 54 aprlnerilsslr--rsvaahpwhdleigp----gapqifnvvdriily 96

RESULT 5
AAM41239
ID AAM41239 standard; Protein; 201 AA.
XX AC AAM41239;
XX AC AAM41239;
DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6170.
XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX OS Homo sapiens.
XX PN W0200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60395.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
```

XX  
PT  
PT

PT treating inflammatory and autoimmune disorders -  
 XX Claim 1; Page 17; 97pp; English.

CC The present sequence is that of mouse interleukin-B60 (IL-B60), a  
 CC novel, small soluble cytokine-like protein that exhibits structural  
 CC motifs characteristic of a member of the long-chain cytokines, and  
 CC which shows homology to granulocyte colony stimulating factor and  
 CC interleukin-6. IL-60B may have either stimulatory or inhibitory  
 CC effects on haematopoietic cells, including e.g. lymphoid cells,  
 CC such as T-cells, B-cells, natural killer cells, macrophages,  
 CC dendritic cells, haematopoietic progenitors, etc. Methods are  
 CC provided for modulating the physiology or development of a cell or  
 CC tissue culture cells by contacting the cell with an agonist or  
 CC antagonist of IL-B60 or an agonist of antagonist of a complex of  
 CC mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see  
 CC AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological  
 CC factor in motor neuron development and regeneration. IL-60B, its  
 CC agonists and antagonists may be used to treat inflammatory or  
 CC autoimmune disorders and also for drug screening.

XX Sequence 215 AA;

Query Match 18.4%; Score 63; DB 21; Length 215;  
 Best Local Similarity 38.9%; Pred. No. 1.6;  
 Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

Qy 7 SRILERSLSRSRVIANVALVFFHDPDIPGVTLTAD 42  
 Db 37 trylehqlrslagtylnylgppfnepdfnprrlgae 72

RESULT 8

AAE00828  
 ID AAE00828 standard; Protein; 223 AA.

AC AAE00828;

DT 02-JUL-2001 (first entry)

DE Human cardiotrophin-like cytokine (CLC) protein.

KW Human; biologically active complex; haemopoietin receptor; NR6;  
 KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;  
 KW differentiation; cell survival; neurotrophic activity.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Signal\_peptide

FT Protein 28..223

FT Protein /label= Human\_mature\_CLC\_protein

FT Protein /note= "Cardiotrophin-like cytokine"

XX WO200127157-A1.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-AU01216.

XX 08-OCT-1999; 99AU-0003327.

XX 12-MAY-2000; 2000AU-0007489.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;

XX Nakata Y, Hasegawa M;

XX WPI; 2001-281978/29.

XX N-PSDB; AAD04201.

PT New biologically active complex comprising NR6 and  
 PT cardiotrophin-like-cytokine, for facilitating proliferation,  
 PT differentiation and/or survival of a cell -

XX Claim 32; Page 114-115; 123pp; English.

XX The present invention relates to a biologically active complex comprising  
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).  
 CC The complex is useful in the manufacture of a medicament for the  
 CC treatment and/or prophylaxis of a subject, as it is involved in  
 CC facilitating proliferation, differentiation and/or survival of a cell.  
 CC The complex or its components have neurotrophic activity. The present  
 CC sequence is human cardiotrophin-like cytokine (CLC) protein.

XX Sequence 223 AA;

Query Match 18.4%; Score 63; DB 22; Length 223;  
 Best Local Similarity 38.9%; Pred. No. 1.7;  
 Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

Qy 7 SRILERSLSRSRVIANVALVFFHDPDIPGVTLTAD 42

Db 47 trylehqlrslagtylnylgppfnepdfnprrlgae 82

RESULT 9

AAW29715

ID AAW29715 standard; Protein; 225 AA.

XX AAW29715;

XX 09-NOV-1998 (first entry)

XX Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;  
 KW peripheral neuropathy; dystrophy; neural retina degeneration;  
 KW common variable immunodeficiency; CVID; selective IgA deficiency;  
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;  
 KW therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Sig\_peptide

FT Protein 28..225

FT Protein /label= Mat\_protein

XX WO9833922-A1.

XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.

XX 30-JAN-1998; 98US-0016534.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M, Elliot GS, Sarmiento U, Senaldi G;

XX WPI; 1998-437475/37.

XX N-PSDB; AAV47510-11.

XX Newly isolated nucleic acid encoding human or murine neurotrophic  
 factor NNT-1 - useful for treatment of neurological and  
 immunological diseases or inflammation, also as vaccine adjuvant

XX Claim 12; Fig 3; 120pp; English.

DR	N-PSDB; AAV47512.	
XX	Newly isolated nucleic acid encoding human or murine neurotrophic factor NNT-1 - useful for treatment of neurological and immunological diseases or inflammation, also as vaccine adjuvant	
PT		
PT		
XX	Claim 13; Fig 5; 120pp; English.	
PS		
CC	This is the amino acid sequence of a murine neurotrophic factor, designated NNT-1, that is a growth factor for neurons and for B or T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512).	
CC	Human NNT-1 (see AAV29715) is also provided. Vectors and host cells for use in the production of human murine recombinant NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids are delivered in gene therapy vectors.	
CC	Sequence 225 AA;	
XX		
SQ		
Query Match	18.4%;	Score 63; DB 19; Length 225;
Best Local Similarity	38.9%;	Pred. No. 1.7;
Matches 14;	Conservative 5; Mismatches 17; Indels 0; Gaps	
QY	7 SRILERSRSRVIANVALVPFHDPDGPVTLTAD 42	
Db	:      ::       :	
	47 tlyleqlrslagtylnylgppfnepdnprrgae 82	
RESULT 11		
AAW56141		
ID	AAW56141 standard; Protein; 225 AA.	
XX		
XX	AAW56141;	
XX		
DT	13-JUL-1998 (first entry)	
XX		
DE	Amino acid sequence of human neurotrophic factor NNT-1.	
XX		
KK	Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron treatment; neurological disease; degeneration; Parkinson's disease; amyotrophic lateral sclerosis; ALS; Alzheimers's disease; stroke.	
XX		
OS	Homo sapiens.	
XX		
FF	Key Location/Qualifiers	
FT	Peptide 1..27	
FT	/note= "signal peptide"	
FT	Protein 28..225	
FT	/note= "mature protein"	
XX		
PN	US5741772-A.	
XX		
XX		
PD	21-APR-1998.	
PF		
PF	03-FEB-1997; 97US-0792019.	
XX		
PP	03-FEB-1997; 97US-0792019.	
XX		
PA	(AMGE-) AMGEN INC.	
PI	Chang M.	

XX WPI; 1998-260526/23.  
DR N-PSDB; AAV22652.  
XX  
PT Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
PT useful for stimulating growth of motor and sympathetic neurons  
XX  
XX Claim 1; Fig 3; 4lpp; English.  
XX  
XX The present sequence represents a human neurotrophic factor, designated  
CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
CC neurons. The NNT-1 protein is useful in the treatment of neurological  
CC diseases characterised by the degeneration and death of particular  
CC classes of neurons. These diseases specifically include Parkinson's  
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
CC stroke and various degenerative disorders affecting vision.  
XX  
XX Sequence 225 AA;  
SQ

Query Match 18.4%; Score 63; DB 19; Length 225;  
Best Local Similarity 38.9%; Pred. No. 1.7;  
Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;  
XX

QY 7 SRILERSLSRSRVIANVALVPFHPDIPGPTLTAD 42  
DB 47 tryleqlrslagtylnylgppfnepdnprrlgae 82

RESULT 12  
AAW56142  
ID AAW56142 standard; Protein; 225 AA.  
XX  
AC AAW56142;  
XX  
DT 13-JUL-1998 (first entry)  
XX  
DE Amino acid sequence of murine neurotrophic factor NNT-1.  
XX  
KW Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;  
KW treatment; neurological disease; degeneration; Parkinson's disease;  
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..27  
FT /note= "signal peptide"  
FT Protein 28..225  
FT /note= "mature peptide"  
FT  
FT  
XX US5741772-A.  
PN  
XX  
XX 21-APR-1998.  
PD  
XX  
XX 03-FEB-1997; 97US-0792019.  
PF  
XX  
XX 03-FEB-1997; 97US-0792019.  
PR  
XX  
XX (AMGE-) AMGEN INC.  
PA  
XX  
XX Chang M;  
PI  
XX  
XX WPI; 1998-260526/23.  
DR N-PSDB; AAV22654.  
XX  
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -  
PT useful for stimulating growth of motor and sympathetic neurons  
XX  
XX Claim 2; Fig 5; 4lpp; English.  
XX  
XX The present sequence represents a murine neurotrophic factor, designated  
CC NNT-1, which is capable of stimulating growth of motor or sympathetic  
CC

CC neurons. The NNT-1 protein is useful in the treatment of neurological  
CC diseases characterised by the degeneration and death of particular  
CC classes of neurons. These diseases specifically include Parkinson's  
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,  
CC stroke and various degenerative disorders affecting vision.  
XX  
XX Sequence 225 AA;  
SQ

Query Match 18.4%; Score 63; DB 19; Length 225;  
Best Local Similarity 38.9%; Pred. No. 1.7;  
Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;  
XX

QY 7 SRILERSLSRSRVIANVALVPFHPDIPGPTLTAD 42  
DB 47 tryleqlrslagtylnylgppfnepdnprrlgae 82

RESULT 13  
AAW94466  
ID AAW94466 standard; Protein; 225 AA.  
XX  
AC AAW94466;  
XX  
DT 22-APR-1999 (first entry)  
XX  
DE Human cardiotrophin-like cytokine protein.  
XX  
KW Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;  
KW CLC; IL-6; diagnosis; detection; immune system-related disorder;  
KW cancer; cardiac disorder; heart failure; hypertension; cancer;  
KW autoimmune disorder; infection.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..27  
FT /label= signal  
FT Protein 28..225  
FT /label= Cardiotrophin-like\_cytokine  
FT Domain 74..79  
FT /label= CD-I  
FT /note= "conserved domain"  
FT Domain 150..156  
FT /label= CD-II  
FT /note= "conserved domain"  
FT Domain 194..198  
FT /label= CD-III  
FT /note= "conserved domain"  
XX WO9900415-A1.  
PN  
XX  
XX 07-JAN-1999.  
PD  
XX  
XX 29-JUN-1998; 98WO-US13129.  
PF  
XX  
XX 30-JUN-1997; 97US-0051311.  
PR  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Ruben SM, Shi Y;  
PI  
XX  
XX WPI; 1999-095678/08.  
DR N-PSDB; AAX16161.  
XX  
XX New isolated cardiotrophin-like cytokine nucleic acid - used to  
PT develop products for treating cardiac and immune system disorders,  
PT e.g. heart failure, hypertension, cancers, autoimmune disorders and  
PT infections  
XX  
XX Claim 1; Fig 1; 103pp; English.  
PS  
XX The present invention relates to a novel cardiotrophin-like cytokine  
CC

(CLC) protein which is a member of the interleukin 6 (IL-6) cytokine family. The present sequence represents the human CLC protein. The present invention also describes screening methods for identifying agonists and antagonists of CLC activity, as well as methods for detecting cardiac and immune system-related disorders and therapeutic methods for treating cardiac and immune system-related disorders, e.g. heart failure, hypertension, cancers, autoimmune disorders and infections.

AA	Sequence	225 AA;
SQ		

Query Match 18.4%; Score 63; DB 20; Length 235;  
Best Local Similarity 38.9%; Pred. No. 1.7;  
Matches 14; Conservative 5; Mismatches 17; Indels

QY	7	SRILERSRSRVANYALVPFHDPDGPVLTAD	42
		:          :	
Db	47	tryleqlrslagtylnylgppfnepdfnpprlgae	82

RESULT 14

AY87813  
ID AY87813 standard; Protein; 225 AA.

AC AAY87813;

DT 24-AUG-2000 (first entry)

Human NNT-1 protein.

NNT-1; human; neurotrophic factor; nontropic; neuroprotective; treatment; anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological; nervous system degeneration; Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; Huntington's disease; peripheral neuropathy; neural retina degeneration; retinopathy; immune disorder; hematopoietic disorder.

OS Homo sapiens.

PN US6054294-A.

25-APR-2000.

12-DEC-1997;

03-FEB-1997: 0705-0702010

XX  
PA (AMGE-) AMGEN INC

XX  
BT  
Cheng M.

XX  
XX  
+  
C  
C  
C

100

N-PSDB; AAA39481.

New nucleic acids

growth of motor or sympathetic neurons for treating neuron cell damage.

Claim 1c; Fig 3; 42pp; English.

This invention describes a novel

novel neurotrophic factor (NNT-1)(III) which has neurotropic, neuroprotective, anticonvulsant, antiparkinsonian, antididiabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic

disorders, and/or dystrophies or degeneration of the neural retina such as retinitis pigmentosa, drug-induced retinopathies, stationary forms of night blindness, progressive cone-rod degeneration, immune disorders and hematopoietic disorders. (I) is effective in treating neurological conditions and promotes neuron regeneration. Neural functions are effectively restored in patients suffering from various neurological disorders. This sequence represents the human NNT-1 protein described in the method of the invention.

AA	Sequence	225 AA;
SQ		

Query Match 18.4%; Score 63; DB 21; Length 225;  
Best Local Similarity 38.9%; Pred. No. 1.7;  
Matches 14; Conservative 5; Mismatches 17; Indels

Qy 7 SRILERSLSRRVIANYALVPFHDPDIPVTLTAD 42  
: : : : :  
Db 47 trylehqrlslagtylnylgppfnepdnprlga 82

RESULT 15

AAV87814  
ID AAV87814 standard; Protein; 225 AA.

AC AAY87814;

24-AUG-2000 (first entry)

XX  
DE  
Murine NNT-1 protein.

XX NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;  
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;  
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;  
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;  
 KW retinopathy; immune disorder; hematopoietic disorder.

Mus sp.

US6054294-A.

XX  
PD  
25-APR-2000.

12-DEC-1997. 07US-00999910

03-SEP-1007 03:00 03:00 03:00

XX

XX  
(page)XX  
Chang M;

WPI; 2000-338492/29.  
N-PSDB: AAA39483

New nucleic acids

growth of motor or sympathetic neurons for treating neuron cell damage

XIX  
 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 25

This invention describes a novel nucleic acid molecule (I) encoding a novel neurotrophic factor (NNT-1) (II) which has neurotrophic, neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic

CC disorders, and/or dystrophies or degeneration of the neural retina such  
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of  
 CC night blindness, progressive cone-rod degeneration, immune disorders and  
 CC hematopoietic disorders. (I) is effective in treating neurological  
 CC conditions and promotes neuron regeneration. Neural functions are  
 CC effectively restored in patients suffering from various neurological  
 CC disorders. This sequence represents the murine NNT-1 protein described in  
 CC the method of the invention.  
 XX  
 SQ Sequence 225 AA;

Query Match 18.4%; Score 63; DB 21; Length 225;  
 Best Local Similarity 38.9%; Pred. No. 1.7;  
 Matches 14; Conservative 5; Mismatches 17; Indels 0; Gaps 0;

QY 7 SRILERSRSRVIANVALVPFHPDIPVTLTAD 42  
 : | | | : | | : | | : | | :  
 Db 47 trylehqlrslagtylnylgpptnepdnprrgae 82

Search completed: March 7, 2002, 22:05:21  
 Job time: 4639 sec





Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	175	51.2	5175	2	T20992	hypothetical prote
2	175	51.2	5198	2	T43290	hemiprotein precurs
3	71.5	20.9	1292	2	T48007	P-glycoprotein hom
4	70.5	20.6	1286	2	T03187	probable ABC trans
5	65	19.0	269	2	S53116	hypothetical prote
6	63.5	18.6	237	1	R5D02	ribosomal protein
7	62.5	18.3	843	2	A27131	epidermal growth f
8	62.5	18.3	1330	1	G0FFE	epidermal growth f
9	62	18.1	1229	2	D55023	P-glycoprotein-lik
10	62	18.1	1229	2	T52319	P-glycoprotein-lik
11	61	17.8	1230	2	E85023	probable P-glycopr
12	60.5	17.7	67	2	D70929	probable ferredoxi
13	60	17.5	211	2	T07594	probable inorganic
14	60	17.5	216	2	T01945	inorganic pyrophos
15	60	17.5	230	2	C86141	protein T25K16.5 [
16	59.5	17.4	178	2	F69716	transcription regu
17	59	17.3	214	2	T01370	inorganic pyrophos
18	59	17.3	552	2	E69438	probable fatty-aci
19	58.5	17.1	275	2	S51658	light harvesting c
20	58.5	17.1	323	2	B48067	ethanolamine-phos
21	58.5	17.1	415	2	B85436	hypothetical prote
22	58.5	17.1	512	2	T43405	probable dihydroli
23	58	17.0	237	2	S75270	hypothetical prote
24	58	17.0	342	2	D49348	succinoglycan bios
25	58	17.0	380	2	C83719	hypothetical prote
26	57	16.7	114	2	T50971	hypothetical prote
27	57	16.7	283	2	T47174	hypothetical prote
28	57	16.7	359	2	T30382	hypothetical prote
29	57	16.7	399	2	B70345	hypothetical prote



Query Match	18.6%	Score 63.5;	DB 1;	Length 237;
Best Local Similarity	36.2%;	Pred. NO. 2.1;		
Matches 21;	Conservative	6;	Mismatches 22;	Indels 9;
				Gaps 3;

C:Genetics:  
A:Cross-references: EMBL:AB09214; NID:g129/019; FIDN:CAA01/00.1; FID:el38/56; FID:g129/00.1  
A:Map position: 12p

A:Title: Characterization of genes which are deactivated upon the onset of development I  
A:Reference number: S09166; MJID:87106347  
A:Contents: annotation; differential expression  
C:Genetics:  
A: Introns: 6/2  
C: Superfamily: Escherichia coli ribosomal protein L2  
C: Keywords: protein biosynthesis; ribosome

Query Match	18.6%	Score 63.5;	DB 1;	Length 237;
Best Local Similarity	36.2%	Pred. NO. 2.1;		
Matches 21;	Conservative	6;	Mismatches 22;	Indels 9;
				Gaps 3;

Matches 18; Conservative 8; Mismatches 14; Indels 13; Gaps 54

F:843/Active site: Lys #status predicted  
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

QY 49 ELRQ--LYVQGGDCPEMSV 66

T07594

probable inorganic pyrophosphatase (EC 3.6.1.1), soluble PPA - potato  
C:Species: Solanum tuberosum (potato)  
C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 11-May-2000  
C:Accession: T07594

R:du Jardin, P.; Rojas-Beltran, J.; Gebhardt, C.; Brasseur, R.  
Plant Physiol. 109, 853-860, 1995  
A:Title: Molecular cloning and characterization of a soluble inorganic pyrophosphatase  
A:Reference number: 216037; MUID:96163189

A:Accession: T07594

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-211 <DUJ>

A:Cross-references: EMBL:Z36894; NID:g534915; PIDN:CAA85362.1; PID:g534916

C:Genetics:

A:Gene: ppa

C:Function:

A:Description: catalyzes the hydrolysis of pyrophosphate

C:Superfamily: inorganic pyrophosphatase

C:Keywords: hydrolase

Query Match 17.5%; Score 60; DB 2; Length 211;  
Best Local Similarity 46.3%; Pred. No. 5.1;  
Matches 19; Conservative 4; Mismatches 12; Indels 6; Gaps 2;

QY 6 ASRLERSLSRSRVIANVALVPFHPDGPVTLTADPVWF 46

DB 14 APRLNERILSSLSR--RSVAHPWHDLEIGPEA----PSVF 48

#### RESULT 14

T01946

Inorganic pyrophosphatase homolog Fl104.12 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 22-Jun-1999

C:Accession: T01946

R:Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevasakis, E.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of A. thaliana Fl104.

A:Reference number: Z14466

A:Accession: T01946

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-216 <ABU>

A:Cross-references: EMBL:AF096370; NID:g3695372; PIDN:AAC62786.1; PID:g3695383

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 46/1; 55/3; 77/3; 113/3; 135/3; 176/1; 203/2

A:Note: Fl104.12

C:Superfamily: inorganic pyrophosphatase

Query Match 17.5%; Score 60; DB 2; Length 216;  
Best Local Similarity 43.6%; Pred. No. 5.2;  
Matches 17; Conservative 6; Mismatches 10; Indels 6; Gaps 2;

QY 8 RILERSLSRSRVIANVALVPFHPDGPVTLTADPVWF 46

DB 21 RLNERILSSLSK--RSVAHPWHDLEIGP----GAPVIF 53

#### RESULT 15

C86141

protein T25K16.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86141

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huijzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Ma  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; T  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86141

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <STO>

A:Cross-references: GB:AE005172; NID:g6715648; PIDN:AAF26475.1; GSPDB:GN00141

C:Genetics:

A:Gene: T25K16.5

A:Map position: 1

Query Match 17.5%; Score 60; DB 2; Length 230;  
Best Local Similarity 51.6%; Pred. No. 5.6;  
Matches 16; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

QY 6 ASRLERSLSRSRVIANVALVPFHPDGP 36

DB 15 APRLNERILSSLSR--RSVAHPWHDLEIGP 43

Search completed: March 7, 2002, 22:06:22

Job time: 3535 sec





DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (EGFR)  
 DE (GURKEN RECEPTOR) (TORPEDO PROTEIN) (DROSOPHILA RELATIVE OF ERBB).  
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyrididae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schubach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in  
 RT subdomains of the receptor protein.";  
 RL Genetics 137:531-550(1994).  
 RN [2]  
 RN REVISIONS.  
 RA Clifford R., Schubach T.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: Conservation of both  
 RT hormone binding and kinase domains.";  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=OREGON-R; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts.";  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II), AND CHARACTERIZATION.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";  
 RL Dev. Biol. 205:129-144(1999).  
 RN [6]  
 RN SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Hoyt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan D.A., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam M.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RN SEQUENCE OF 959-1078 FROM N.A.  
 RP STRAIN=DAEKWANYEONG;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;  
 RT "A Drosophila genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RN PARTIAL SEQUENCE FROM N.A., AND MUTATION ANALYSIS.  
 RP MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.-Z.;  
 RT "Interallelic complementation among DER/flb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RN REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "There must be 50 ways to rule the signal: the case of the Drosophila  
 RL EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -|- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS.  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -|- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -|- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -|- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -|- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GLANDS.  
 CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF052754; AAC08536.1; -;  
 CC EMBL; AF052753; AAC08536.1; JOINED.  
 CC EMBL; AF052754; AAC08535.1; -;  
 CC EMBL; AF052752; AAC08535.1; JOINED.  
 CC EMBL; K03054; AAA51462.1; -;  
 CC EMBL; K03417; AAA51460.1; -;





CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Y13803; CAA74130.1; -;  
DR EMBL; AL583925; CAC31918.1; -;  
DR HSP; P14604; 2DUB.  
DR Leproma; ML2402; -;  
DR InterPro; IPR001753; Enoyl-CoA\_hydratase.  
DR Pfam; PF00378; ECH; 1.  
DR PROSITE; PS00166; Enoyl-CoA-HYDRATASE; 1.  
KW Fatty acid metabolism; Lyase; Complete proteome.  
SQ SEQUENCE 257 AA; 27516 MW; 25B801EB690BCD00 CRC64;

Query Match 17.7%; Score 60.5; DB 1; Length 257;  
Best Local Similarity 27.0%; Pred. No. 2.7;  
Matches 20; Conservative 7; Mismatches 28; Indels 19; Gaps 2;

QY 2 VIDGASRI-----LERSLSRSRVIANVALVPHDPDIPGVTLTADPVVQREL 51  
DB 7 LVGGQVRGVIITLNRPOALNALSQMMNEITNAKELDIDPDVGAILTSPKVF----- 61  
QY 52 QLYVQGGDCPEMS 65  
DB 62 -----AAGADIKEMA 71

RESULT 5  
ID IPYR\_SOLTU STANDARD; PRT; 211 AA.  
AC Q43187;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SOLUBLE INORGANIC PYROPHOSPHATASE (EC 3.6.1.1) (PYROPHOSPHATE PHOSPHO-  
DE HYDROLASE) (PPASE).  
GN PPA.  
OS Solanum tuberosum (Potato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=1113;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=96163189; PubMed=8552717;  
RA du Jardin P., Rojas-Beltran J., Gebhardt C., Brasseur R.;  
RT "Molecular cloning and characterization of a soluble inorganic  
RT pyrophosphatase in potato."  
RL Plant Physiol. 109:853-860(1995).  
CC -1- CATALYTIC ACTIVITY: PYROPHOSPHATE + H(2)O = 2 ORTHOPHOSPHATE.  
CC -1- COFACTOR: MAGNESIUM DEPENDENT.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE PPASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z36894; CAA85362.1; -;  
DR HSP; P38576; 2PRD.  
DR InterPro; IPR001596; Pyrophosphatase.  
DR Pfam; PF00719; Pyrophosphatase; 1.  
DR ProDom; PD002014; Pyrophosphatase; 1.  
DR PROSITE; PS00387; PPASE; 1.  
KW Hydrolase; Magnesium.  
FT ACT\_SITE 61  
SQ SEQUENCE 211 AA; 24261 MW; D06115FC6F2AC22A CRC64;

BY SIMILARITY.

Query Match 17.5%; Score 60; DB 1; Length 211;  
Best Local Similarity 46.3%; Pred. No. 2.5;  
Matches 19; Conservative 4; Mismatches 12; Indels 6; Gaps 2;

QY 6 ASRILERSLSRSRVIANVALVPHDPDIPGVTLTADPVV 46  
DB 14 APRLNERILSSISR--RSVAHPWHDLGPEA-----PSVF 48

RESULT 6  
ID SPST\_BACSU STANDARD; PRT; 178 AA.  
AC P37554;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE STAGE V SPOULATION PROTEIN T.  
GN SPOVT.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin."  
RL DNA Res. 1:1-14(1994).  
RN [2]

RP FUNCTION.  
RX MEDLINE=96345614; PubMed=8755877;  
RA Bayan I., Hobot J., Cutting S.M.;  
RT "A compartmentalized regulator of developmental gene expression in  
RT Bacillus subtilis."  
RL J. Bacteriol. 178:4500-4507(1996).  
CC -1- FUNCTION: POSITIVE AND NEGATIVE TRANSCRIPTIONAL REGULATOR OF SIGMA  
CC G-DEPENDENT GENES. MAY PROVIDE A MECHANISM OF FEEDBACK CONTROL  
CC THAT IS IMPORTANT FOR FORESPORE DEVELOPMENT.  
CC -1- SIMILARITY: STRONG, TO B.SUBTILIS ABRB AND ABH. SOME, TO  
CC M.TUBERCULOSIS RV2595.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D26185; BAA05291.1; -;  
DR EMBL; Z99104; CAB11832.1; -;  
DR Subtilist; BG10119; spoVT.  
KW Transcription regulation; DNA-binding; Repressor; Activator;  
KW Sporulation; Complete proteome.  
SQ SEQUENCE 178 AA; 19742 MW; 19418022DD371180 CRC64;

Query Match 17.48%; Score 59.5; DB 1; Length 178;  
Best Local Similarity 32.1%; Pred. No. 2.4;  
Matches 18; Conservative 10; Mismatches 15; Indels 13; Gaps 3;

QY 7 SRILERSLSRSRVIANVA---LVPHDPD-----IGPVTLTADP---VVFORE 49  
DB 99 SEMLERTMDQRSSVLESDAKSVQLVNGIDEDMNSYTVGPVANGDPGAVVIFSKD 154

RESULT 7  
ID IPYR\_MAIZE STANDARD; PRT; 214 AA.  
AC O48556;



[2]  
SEQUENCE FROM N.A.  
RC STRAIN=RCR2011 / SU47;  
RX MEDLINE=94162682; PubMed=8118055;  
RA Becker A., Kleickmann A., Kuester H., Keller M., Arnold W.,  
RA Puehler A.;  
RT Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT, and  
RT exoU involved in exopolysaccharide biosynthesis and nodule invasion:  
RT exoU and exoW probably encode glycosyltransferases.";  
RL Mol. Plant Microbe Interact. 6:735-744(1993).  
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF  
CC SUCCINOGLYCAN (EPS I) NEEDED FOR THE ADDITION OF THE SIXTH SUGAR  
CC (GLUCOSE), CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN  
CC THE FIFTH AND SIXTH SUGAR.  
CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L20758; AAA16053.1; -;  
CC EMBL: Z22646; CAA80359.1; -;  
CC PIR: D49348; D49348.  
CC InterPro: IPR001173; Glycos.transf.2.  
CC Pfam: PF00535; Glycos.transf.2.1.  
CC Transferase; Glycosyltransferase; Exopolysaccharide synthesis;  
CC Plasmid  
CC SEQUENCE 342 AA; 37017 MW; EAF550EBDA023BC CRC64;  
SQ

	Query Match	17.0%; Score 58; DB 1; Length 342;
	Best Local Similarity	37.3%; Pred. No. 7.7;
	Matches 22; Conservative	5; Mismatches 28; Indels 4; Gaps 3;
QY	4 DGASRI-LERSLSRSRVIA-NYALVFPFHDPDIPGVTLTADPPVFQRELRQLYYOGGD	60
DB	58 DGTGRLVNVRFEENRGPAAARNHAIATSHSPILIG--VLDAADFFPFGRGLQSQQDGWD	114
	: : :	:
RESULT 10		
ID GRK_BACHD	STANDARD; PRT; 380 AA.	
AC Q9Z9P2; Q9KFC9;		
DT 20-AUG-2001 (Rel. 40, Created)		
DT 20-AUG-2001 (Rel. 40, Last sequence update)		
DE 20-AUG-2001 (Rel. 40, Last annotation update)		
DE GLYCERATE KINASE (EC 2.7.1.31).		
GN GLXK OR BH0555.		
OS Bacillus halodurans.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC Bacillus/Staphylococcus group; Bacillus.		
OX NCBI_TaxID=86665;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=C-125 / JCM 9153;		
RX MEDLINE=20512582; PubMed=11058132;		
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA Fuji F., Hirama C., Nakamura Y., Ogawara S.,		
RA Horikoshi K.;		
"Complete genome sequence of the alkaliphilic bacterium Bacillus		
RT halodurans and genomic sequence comparison with Bacillus subtilis.";		
RL Nucleic Acids Res. 28:4317-4331(2000).		
RN [2]		
RP SEQUENCE OF I-189 FROM N.A.		
RR STRAIN=C-125 / JCM 9153;		
RX MEDLINE=99184645; PubMed=10086841;		
RA Takami H., Nakasone K., Hirama C., Takaki Y., Masui N., Fujii F.,		

```

RA Nakamura Y., Inoue A.;
RT "An improved physical and genetic map of the genome of alkaliphilic
RL Bacillus sp. C-125.";
RL Extremophiles 3:21-28(1999).
CC -!- CATALYTIC ACTIVITY: ATP + (R)-GLYCERATE -> ADP + 3-PHOSPHO-(R)-
CC GLYCERATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001508; BAB04274.1; -
DR EMBL; AB013375; BAA75390.1; -
DR InterPro; IPR003747; DUF168.
DR Pfam; PF02595; DUF168; 1.
KW Transferase; Kinase; Complete proteome.
SQ SEQUENCE 380 AA; 39955 MW; 0C6E227253B3E46D CRC64;

Query Match 17.0%; Score 58; DB 1; Length 380;
Best Local Similarity 35.4%; Pred. No. 8.7;
Matches 23; Conservative 7; Mismatches 27; Indels 8; Gaps 4;

QY 6 ASRLERSLSRSRVI--ANYALVPFHDPDIPVTLTADPVVFQRELRLQIYVGG- GDGP 62
| : | : | | | | | | | | | | | | | | | | | | | | | | | |
DB 20 AARAEIKGF---RRVIPNANVRLIPMADGGEGTVQSIVD--ALQGERKKVVKVEPLGDTV 74
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 63 EMSVG 67
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 75 EAEYG 79
| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 11
FDH_PICAN
ID FDH_PICAN STANDARD; PRT; 361 AA.
AC P33677;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FORMATE DEHYDROGENASE (EC 1.2.1.2) (NAD-DEPENDENT FORMATE
DE DEHYDROGENASE) (FDH).
GN FMDH.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RR Hollenberg C.P., Janowicz Z.;
RT "DNA-molecules coding for FMDH control regions and structured gene for
RT a protein having FMDH-activity and their uses.";
RL Patent number EP0299108, 18-JAN-1989.
SC -1- CATALYTIC ACTIVITY: FORMATE + NAD(+) = CO(2) + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY. STRONG, TO OTHER FDH.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; A06214; CAA00531.1; -.
CC HSSP; P33160; 2NAC.
DR InterPro: IPR002162; D_2_hydroxyacid_DH.
DR Pfam: PF00389; 2-Hacid_DH; 1.

```

```

DR PROSITE: PS00065; D_2_HYDROXYACID_DH.1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH.2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH.3; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0 0 PROBABLE.
FT NP_BIND 164 199 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 257 257 SUBSTRATE-BINDING (BY SIMILARITY).
SQ SEQUENCE 361 AA; 39779 MW; 0FC001366F9E479B CRC64;

Query Match
Best Local Similarity 16.8%; Score 57.5; DB 1; Length 361;
Matches 15; Conservative 14; Mismatches 18; Indels 9; Gaps 2;

QY 4 DGASRIERSLSRSRVIANVALVPFHPDIPGVTITADVPVVFQRELRLQLYVGGG 59
Db 47 EGQSVLEKNISDADVIIS---TPFH-----PAYITKERIDAKKLLKLLVAGVG 93

RESULT 12
Y545_AQUAE
ID Y545_AQUAE STANDARD; PRT; 399 AA.
AC O66822;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_545.
GN AQ_545.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
ON NCBI_TaxId=63363;
RX STRAIN=VF5;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000695; AAC06792.1; -.
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 399 AA; 46953 MW; AA57D8680674318B CRC64;

Query Match
Best Local Similarity 16.7%; Score 57; DB 1; Length 399;
Matches 13; Conservative 11; Mismatches 9; Indels 10; Gaps 2;

QY 4 DGASRIERSLSRSRVIANVALVPFHPDIPGVTITADVPVVF 46
Db 150 DG-EKVMKALVDVRKVLADYE-----GPVVITEDKVIY 182

RESULT 13
COIL_MOUSE
ID COIL_MOUSE STANDARD; PRT; 484 AA.
AC Q9NUM3; Q9CVA2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CORONIN 1B (CORONIN 2).
GN CORO1B.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RN [2]
SQ SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
RT humans and mice: close genetic linkage between coronin-2 and CD45-
RT associated protein.";
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RN [2]
SQ SEQUENCE OF 163-484 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Stomach;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: MAY BE INVOLVED IN CYTOKINESIS, MOTILITY, AND SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF143956; AAD32704.1; -.
CC EMBL: AK008947; BAB25985.1; -.
CC MGD; MGI:1345963; Coro1b.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 3.
CC SMART; SM00320; WD40; 3.
CC PROSITE; PS00678; WD_REPEATS_1; 1.
CC PROSITE; PS50082; WD_REPEATS_2; 2.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Actin-binding; Repeat; WD repeat; Coiled coil.
KW REPEAT 80 120 WD 1.
FT REPEAT 130 170 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 217 260 WD 4.
FT REPEAT 265 305 WD 5.
FT REPEAT 444 482 COILED COIL (POTENTIAL).
FT DOMAIN 393 393 R -> G (IN REF. 2).
FT CONFLICT 393 393
SQ SEQUENCE 484 AA; 53912 MW; 9631CC02E7EAC72F CRC64;

Query Match
Best Local Similarity 16.7%; Score 57; DB 1; Length 484;
Best Local Similarity 31.4%; Pred. No. 15;

```

Query Match 16.5%; Score 56.5; DB 1; Length 1094;  
Best Local Similarity 33.3%; Pred. No. 45;  
Matches 17; Conservative 9; Mismatches 16; Indels 9; Gaps 3;  
QY 25 ALVPFHDPDGP--VLTADPVVFQRLRLQYQ---GGGD----CPMSV 66

Db 882 ALLPHMDTPLGPAAAAALTRQIGVRDLRELDFFEMPLAGGDLRGSPDVS 932

Search completed: March 7, 2002, 22:13:06  
Job time: 459 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 22:07:47 ; Search time 65.89 Seconds  
(without alignments)  
148.736 Million cell updates/sec

Title: US-09-665-728-1\_COPY\_56\_122

Perfect score: 342

Sequence: 1 QVIDGASRILERSLSRSRV.....RELRLYVQGGDCPMSVG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL\_17.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mmc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	51.2	5198	5 076518	076518 caenorhabdi
2	131.5	38.5	873	11 092108	Q92108 mus musculu
3	131.5	38.5	891	11 09JHA8	Q9JHA8 mus musculu
4	131	38.3	536	4 Q9UMP9	Q9UMP9 homo sapien
5	131	38.3	852	4 Q9Y334	Q9Y334 homo sapien
6	71.5	20.9	1292	10 Q9M109	Q9M109 arabidopsis
7	70.5	20.6	1286	10 080725	080725 arabidopsis
8	65.5	19.2	584	2 Q9XC77	Q9XC77 streptomyce
9	65	19.0	269	3 Q12395	Q12395 saccharomyc
10	63	18.4	225	4 Q9UBD9	Q9UBD9 homo sapien
11	63	18.4	225	11 Q9QZM3	Q9QZM3 mus musculu
12	63	18.4	744	2 Q06693	Q06693 vibrio mimi
13	62.5	18.3	687	2 Q9ABL3	Q9ABL3 caulobacter
14	62.5	18.3	3487	2 Q9FBU0	Q9FBU0 microcystis
15	62	18.1	1229	10 Q49749	Q49749 arabidopsis
16	62	18.1	1229	10 Q9SY12	Q9SY12 arabidopsis
17	61.5	18.0	1392	12 Q9QTC5	Q9QTC5 gallid herp
18	61	17.8	802	3 Q9HE60	Q9HE60 neurospora
19	61	17.8	1230	10 Q9SY13	Q9SY13 arabidopsis

20	60.5	17.7	67	2 053937	053937 mycobacteri
21	60.5	17.7	285	4 099773	099773 homo sapien
22	60.5	17.7	487	12 09QOM9	09QOM9 pseudorabdi
23	60.5	17.7	669	5 09VR40	09VR40 drosophila
24	60	17.5	216	10 082597	082597 arabidopsis
25	60	17.5	230	10 09MAM9	09MAM9 arabidopsis
26	60	17.5	290	11 09DL18	09DL18 mus musculu
27	59.5	17.4	380	3 093982	093982 zygocacchar
28	59.5	17.4	385	5 09NA54	09NA54 caenorhabdi
29	59.5	17.4	443	4 09HAL4	09HAL4 homo sapien
30	59.5	17.4	702	4 09HOK9	09HOK9 homo sapien
31	59	17.3	130	5 09G252	09G252 tainarys so
32	59	17.3	215	10 09SW10	09SW10 populus tre
33	59	17.3	365	2 09KH85	09KH85 streptomyce
34	59	17.3	552	1 028762	028762 archaeoglob
35	58.5	17.1	275	10 040771	040771 picea abies
36	58.5	17.1	354	3 09UOV5	09UOV5 ustilago ma
37	58.5	17.1	373	10 09FUA4	09FUA4 arabidopsis
38	58.5	17.1	415	10 023192	023192 arabidopsis
39	58	17.0	237	2 073158	073158 synechocyst
40	58	17.0	511	4 09H401	09H401 homo sapien
41	57.5	16.8	1820	5 09VVG0	09VVG0 drosophila
42	57	16.7	114	3 09P3Q9	09P3Q9 neurospora
43	57	16.7	192	2 09RMP5	09RMP5 xenorhabdus
44	57	16.7	283	4 09NSK3	09NSK3 homo sapien
45	57	16.7	359	12 09YMT9	09YMT9 lymantria d

## ALIGNMENTS

RESULT 1

076518	076518	PRELIMINARY;	PRT; 5198 AA.
ID	076518	Q10036;	
AC	076518	01-NOV-1998 (TReMBLrel. 08, Created)	
DT	01-NOV-1998	(TReMBLrel. 08, Last sequence update)	
DT	01-JUN-2001	(TReMBLrel. 17, Last annotation update)	
DE	HEMICENTIN PRECURSOR.		
GN	HM-4 OR F15G9.4.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-BRISTOL N2;		
RA	Vogel B.E., Hedgecock E.M.;		
RT	"Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and Germ-Line Chromosome Segregation in C. elegans."		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-BRISTOL N2;		
RA	Sulston J., Kershaw J.;		
RL	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.		
CC	-!- ALTERNATIVE PRODUCTS: TWO FORMS: ISOFORM F15G9.4A AND ISOFORM F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN.		
CC	EMBL; AF074901; AAC26792.1; -		
DR	EMBL; Z47068; CAA87336.1; -		
DR	EMBL; Z47070; CAA87336.1; JOINED.		
DR	EMBL; Z47070; CAA87345.1; -		
DR	EMBL; Z47068; CAA87345.1; JOINED.		
DR	EMBL; Z47068; CAA87335.1; -		
DR	EMBL; Z47070; CAA87335.1; JOINED.		
DR	EMBL; Z47070; CAA87344.1; -		
DR	HSSP; P35555; LEWN.		
DR	WormPep; F15G9.4A; CE18595.		
DR	WormPep; F15G9.4B; CE18596.		
DR	InterPro; IPR000152; Asx_hydroxyl.		

DR InterPro: IPR000515; BPD\_transp.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF Ca.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003600; Ig\_Like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; 1g; 47.  
 DR SMART: SM00179; EGF\_CA; 1.  
 DR SMART: SM00001; EGF\_Like; 1.  
 DR SMART: SM00408; IGC2; 44.  
 DR SMART: SM00410; IGC2; 44.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS00402; BPD\_TRANS\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_CA; 2.  
 KW Alternative splicing; Calcium-binding; EGF-like domain; Glycoprotein;  
 FT Repeat; Signal.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 5198 HEMICENTIN.  
 FT VARSPLIC 1462 5198 MISSING (IN ISOFORM F15G9.4A).  
 SQ SEQUENCE 5198 AA; 570816 MW; DA8511FFB258D37B CRC64;

Query Match 51.2%; Score 175; DB 5; Length 5198;  
 Best Local Similarity 45.5%; Pred. NO. 6.6e-12;  
 Matches 30; Conservative 16; Mismatches 20; Indels 0; Gaps 0;  
 QY 1 QVTDGASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORELRLQLYVGGGD 60  
 DB 53 QVREGAAKIFKTVMAOREKLIYINVPFHPDIPGLIINTDTSTFMKLSKVYVHGGD 112  
 QY 61 CPMSV 66  
 DB 113 CPKTL 118

RESULT 2  
 Q921Q8 PRELIMINARY; PRT; 873 AA.  
 AC Q921Q8;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 94.9 KDA PROTEIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RA Rowen L., Madan A., Oin S., Shaffer T., Ratcliffe A., Abbasi N.,  
 RA Dickhoff R., James R., Loretz C., Lasky S., Hood L.;  
 RT "Sequence of the mouse major histocompatibility locus class III  
 region.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF109905; AAC84152.1;  
 DR InterPro: IPR002035; VWFA.  
 DR SMART: SM00327; VWA; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 873 AA; 94894 MW; 5761E3620F8D831C CRC64;

Query Match 38.5%; Score 131.5; DB 11; Length 873;  
 Best Local Similarity 47.5%; Pred. NO. 1.7e-07;  
 Matches 28; Conservative 9; Mismatches 21; Indels 1; Gaps 1;  
 QY 6 ASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORELRLQLYVGGGDPEM 64  
 DB 346 ARRIVEQRQSGPMEPVF-YILVFPHPDIPGVFTTSDPDSFQKLNEIHALGGGDEPM 403

RESULT 3  
 Q9JHA8 PRELIMINARY; PRT; 891 AA.  
 AC Q9JHA8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE G7C (G7C PROTEIN).  
 GN G7C.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/RIJ, AND BALB/C; TISSUE=BRAIN;  
 RX MEDLINE=98211706; PubMed=9551980;  
 RA Snoek M., Teuscher C., van Vugt H.;  
 RT "Molecular analysis of the major MHC recombinational hot spot located  
 RT within the G7c gene of the murine class III region that is involved in  
 RT disease susceptibility.";  
 RL J. Immunol. 160:266-272(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/RIJ, AND BALB/C; TISSUE=BRAIN;  
 RX MEDLINE=20260998; PubMed=10803853;  
 RA Snoek M., Albertella M.R., van Kooij M., Wixon J., van Vugt H.,  
 RA de Groot K., Campbell R.D.;  
 RT "G7c, a novel gene in the mouse and human major histocompatibility  
 RT complex class III region, possibly controlling lung tumor  
 RT susceptibility.";  
 RL Immunogenetics 51:383-386(2000).  
 DR EMBL; AF134318; AAF69177.1;  
 DR EMBL; AF134319; AAF61401.1;  
 DR InterPro: IPR002035; VWFA.  
 DR SMART: SM00327; VWA; 1.  
 SQ SEQUENCE 891 AA; 95983 MW; 1EA26B5121118720 CRC64;

Query Match 38.5%; Score 131.5; DB 11; Length 891;  
 Best Local Similarity 47.5%; Pred. NO. 1.7e-07;  
 Matches 28; Conservative 9; Mismatches 21; Indels 1; Gaps 1;  
 QY 6 ASRLERSLSRSRVIANVLPFHPDIPGVTLTADPVVFORELRLQLYVGGGDPEM 64  
 DB 338 ARRIVEQRQSGPMEPVF-YILVFPHPDIPGVFTTSDPDSFQKLNEIHALGGGDEPM 395  
 RESULT 4  
 Q9UMP9 PRELIMINARY; PRT; 536 AA.  
 AC Q9UMP9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE G7C PROTEIN (FRAGMENT).  
 GN G7C.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aguado B.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Albertella M.R., Campbell R.D.;  
 RT "Characterisation of the novel gene G7c located in the class III  
 RT region of the human Major Histocompatibility Complex.";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=96411681; PubMed=8812450;  
 RA Albertella M.A.; Jones H.; Thomson W.; Olavesen M.G.; Campbell R.D.;  
 RT "Localization of eight additional genes in the human major  
 RT histocompatibility complex, including the gene encoding the casein  
 RT kinase II beta subunit (CSNK2B).";  
 RL Genomics 36:240-251(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Albertella M.A.;  
 RL Thesis (1997), University of Oxford, Department of Biochemistry,  
 RL MRC Immunochimistry Unit.  
 DR EMBL; AJ245418; CAB52192.1; -  
 FT NON\_TER 1  
 FT NON\_TER 536  
 SQ SEQUENCE 536 AA; 57837 MW; B32D8DD8E24BD06D CRC64;

Query Match 38.3%; Score 131; DB 4; Length 536;  
 Best Local Similarity 44.4%; Pred. No. 1.1e-07;  
 Matches 28; Conservative 10; Mismatches 21; Indels 4; Gaps 1;

QY 6 ASRIERLSRSR-----VIANVALVPFHPDIPGVTITADPVVQRELRLYVGGGDC 61  
 DB 34 AAKIQARHLVEQRGSPMEPVHYLVLPFHPGFGVFTSDPSDFWQQLNEIHALGGDE 93  
 QY 62 PEM 64  
 DB 94 PEM 96

RESULT 5  
 ID Q9V334 PRELIMINARY; PRT; 852 AA.  
 AC Q9V334;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE NG37.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L.; Qin S.; Madan A.; Dickhoff R.; Dors M.; Madan A.; Hicks P.;  
 RA Loretz C.; Ratcliffe A.; Abbasi N.; Shaffer T.; Hood L.;  
 RT "Sequence of the human major histocompatibility complex class III  
 RT region.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF134726; AAD21820.1; -  
 SQ SEQUENCE 852 AA; 92538 MW; 025430B3F912A941 CRC64;

Query Match 38.3%; Score 131; DB 4; Length 852;  
 Best Local Similarity 44.4%; Pred. No. 1.9e-07;  
 Matches 28; Conservative 10; Mismatches 21; Indels 4; Gaps 1;

QY 6 ASRIERLSRSR-----VIANVALVPFHPDIPGVTITADPVVQRELRLYVGGGDC 61  
 DB 332 AAKIQARHLVEQRGSPMEPVHYLVLPFHPGFGVFTSDPSDFWQQLNEIHALGGDE 391  
 QY 62 PEM 64  
 DB 392 PEM 394

RESULT 6  
 ID Q9MIQ9 PRELIMINARY; PRT; 1292 AA.  
 AC Q9MIQ9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE P-GLYCOPROTEIN-LIKE PROTEIN.  
 GN T17J13.110.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 RN NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M.; Mueller-Auer S.; Zipp M.; Schaefer M.; Meves H.W.;  
 RA Lencke K.; Mayer K.F.X.; Queier F.; Salanoubat M.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 FT NON\_TER 1  
 FT NON\_TER 536  
 SQ SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC  
 CC TRANSPORTERS).  
 DR EMBL; AL138651; CAB71875.1; -  
 DR InterPro; IPR003593; AAA.  
 DR InterPro; IPR001140; ABC\_transporter\_tmem.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR InterPro; IPR001687; ATP\_GTP\_A.  
 DR InterPro; IPR001525; C5\_DNA\_meth.  
 DR Pfam; PF00664; ABC\_membrane; 2.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 2.  
 DR PROSITE; PS00095; C5\_MTASE\_2; UNKNOWN\_1.  
 KW ATP-binding; Transport.  
 SQ SEQUENCE 1292 AA; 139753 MW; 9C95DF3AC84E7432 CRC64;

Query Match 20.9%; Score 71.5; DB 10; Length 1292;  
 Best Local Similarity 29.1%; Pred. No. 6;  
 Matches 23; Conservative 10; Mismatches 23; Indels 23; Gaps 3;

QY 9 ILERSLSRSRVIANVALVPFHPDIPGVTITADPVVQRELRLYVGGGDC 48  
 DB 1081 LVGSGSGKSTVIA--LLQRFYDPDQSGITLDGVEIKTLQKWLKRLQOTGLVSGPEVLFNE 1138  
 QY 49 ELRLYVQG--GGDCPEMSV 66  
 DB 1139 TIRANIAYGKGGDATETEI 1157

RESULT 7  
 ID O80725 PRELIMINARY; PRT; 1286 AA.  
 AC O80725;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE PUTATIVE P-GLYCOPROTEIN.  
 GN F14M4.17.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 RN NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RA Rounsley S.D.; Lin X.; Kaul S.; Shea T.P.; Fujii C.Y.; Mason T.M.;  
 RA Shen M.; Ronning C.M.; Fraser C.M.; Somerville C.R.; Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC  
 CC TRANSPORTERS).  
 DR EMBL; AC004411; AAC34225.1; -  
 DR HSSP; P13569; INBD.  
 DR InterPro; IPR003593; AAA.

DR InterPro: IPR001140; ABC\_transporter\_tmem.  
DR InterPro: IPR003439; ABC\_transporter.  
DR InterPro: IPR001687; ATP\_GTP\_A.  
DR InterPro: IPR001525; C5\_DNA\_meth.  
DR Pfam: PF00664; ABC\_membrane; 2.  
DR Pfam: PF00005; ABC\_tran; 2.  
DR SMART: SM00382; AAA; 2.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; 2.  
DR PROSITE: PS00095; C5\_MTASE\_2; UNKNOWN\_1.  
KW ATP-binding; Transport.  
SQ SEQUENCE 1286 AA; 139027 MW; 97D2A8AFBA698E6 CRC64;

Query Match 20.6%; Score 70.5; DB 10; Length 1286;  
Best Local Similarity 27.8%; Pred. No. 7.9;  
Matches 22; Conservative 11; Mismatches 23; Indels 23; Gaps 3;  
  
QY 9 ILERSLSRSRVIANVLPFHPDPDIPVTL-----TADPVVFOR 48  
DB 1075 LVGSGSGKSTVIA--LQRFYDPSGSEITLDGVEIKSLRLKWLKRLQQTGLVLSQEPILFNE 1132  
QY 49 ELRLQVVG-GGDCPEMSV 66  
DB 1133 TIRANTAYKGGDASESEI 1151

RESULT 8  
Q9XCV7 PRELIMINARY; PRT; 584 AA.  
AC Q9XCV7;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE JADJ.  
OS Streptomyces venezuelae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=54571;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISP5230;  
RA Kulowski K., Wendt-Plenkowski E., Han L., Yang K., Vining L.C.,  
Hutchinson C.R.;  
RT "functional characterization of the jadI gene as a cyclase that forms  
anacyclonones.";  
RL J. Am. Chem. Soc. 0:0-0(1999).  
CC -1- COFACTOR: BIOTIN (BY SIMILARITY).  
DR EMBL; AF126429; RAD37851.1; -;  
DR HSSP; P24182; IBNC.  
DR InterPro: IPR001882; Biotin.  
DR InterPro: IPR000089; Biotin\_lipoyl.  
DR InterPro: IPR000901; CPase.  
DR Pfam: PF00364; biotin\_lipoyl; 1.  
DR Pfam: PF00289; CPase\_LChain; 1.  
DR PROSITE: PS00188; BIOTIN; 1.  
DR PROSITE: PS00867; CPASE\_2; UNKNOWN\_1.  
KW Biotin.  
SQ SEQUENCE 584 AA; 61926 MW; 4384224B32ADC64B CRC64;

Query Match 19.2%; Score 65.5; DB 2; Length 584;  
Best Local Similarity 27.8%; Pred. No. 13;  
Matches 22; Conservative 11; Mismatches 19; Indels 27; Gaps 5;  
  
QY 2 VIDGASRLERSLSRSRVIANVLPFHPDPDIPVTLADP----- 43  
DB 387 IVTGASR--EQALQRAARVAERVEGMATAIPFHQAVVTDPD-----FTADPFVRVHTRW 439  
QY 44 --VVFQRELRQLVVGQGD 60  
DB 440 IETEFVNEIKPFAPAGEAD 458

RESULT 9  
Q12395 PRELIMINARY; PRT; 269 AA.  
AC Q12395;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE ORF YLR128W.  
GN YLR128W OR L3111 OR L9233.2.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FY23/RD005;  
RA Verhasselt P., Volckaert G.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RA Delius H.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Delius H., Hebling U.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C (AB972);  
RA Pauley A.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C (AB972);  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
Hawkins J., Hillier L., Jier M., Johnson D.,  
Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
Taich A., Travaaskis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
Wilson R., Waterston R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C (AB972);  
RA Waterston R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X89514; CAA61706.1; -;  
DR EMBL; X91258; CAA62639.1; -;  
DR EMBL; Z73300; CAA97697.1; -;  
DR EMBL; U53877; AAB82374.1; -;  
DR SGD; S0004118; YLR128W.  
SQ SEQUENCE 269 AA; 32203 MW; C8D829C941466180 CRC64;

Query Match 19.0%; Score 65; DB 3; Length 269;  
Best Local Similarity 32.7%; Pred. No. 5.8;  
Matches 16; Conservative 10; Mismatches 19; Indels 4; Gaps 1;  
  
QY 10 LERSLSRSRVIANVLPFHPDPDIPG----PVTLTADPVVFORLRQLY 54  
DB 31 VSKRYLQRNHNINVALNDYDKETGTFDDEVSTVAHPVPKELTQVF 79  
RESULT 10  
Q9UBD9 PRELIMINARY; PRT; 225 AA.  
ID Q9UBD9



049749	ID 049749	PRELIMINARY;	PRT; 1229 AA.
--------	-----------	--------------	---------------

Db 1075 TIRSNIA YKGGDEASEAEI 1094

Search completed: March 7, 2002, 22:07:49  
Job time: 3167 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 20:01:47 ; Search time 148.15 seconds  
(without alignments)  
5966.266 Million cell updates/sec

Title: US-09-665-728-2  
Perfect score: 1031  
Sequence: 1 tctagcgaacccttcggcc.....agcccggtcgcgcgcgcgc 1031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1031	100.0	1031	22	AAF82464
2	51.6	5.0	2004	18	AAF85356
3	50.4	4.9	2338	19	AAV23249
4	50.4	4.9	2338	21	AAZ38195
5	47.2	4.6	2338	12	AAQ14183
6	43.4	4.2	2745	22	AAD09568
7	41.8	4.1	9063	22	AAH23688
8	41.2	4.0	1260	19	AAV41730
9	41	4.0	1747	20	AAV82459
10	40.8	4.0	1160	20	AAZ52508
11	40.4	3.9	2271	15	AAQ62176

12	40.4	3.9	2271	15	AAQ45334
13	40.2	3.9	2214	19	AAV22682
14	40.2	3.9	3331	19	AAV22683
15	40.2	3.9	3331	20	AAZ32021
16	40.2	3.9	3331	22	AAZ32021
17	39.6	3.8	2502	16	AAQ80752
18	39.6	3.8	2502	19	AAV65782
19	39.6	3.8	2502	20	AAV63403
20	39.4	3.8	4748	22	AAQ06573
21	38.8	3.8	1929	19	AAV57472
22	38.8	3.8	2668	21	AAQ03337
23	38.8	3.8	12152	22	AAQ08699
24	38.6	3.7	1598	14	AAQ40461
25	38.4	3.7	2461	17	AAQ14154
26	38.2	3.7	1995	12	AAQ14184
27	38.2	3.7	1995	19	AAV23250
28	38.2	3.7	1995	21	AAZ38196
29	38	3.7	1588	22	AAH26304
30	37.6	3.6	567	21	AAZ29550
31	37.6	3.6	71989	21	AAZ29349
32	37.4	3.6	571	20	AAV89296
33	37.4	3.6	3120	21	AAQ12510
34	37.4	3.6	5145	21	AAQ98196
35	37.4	3.6	6085	18	AAQ70153
36	37.4	3.6	31422	21	AAQ92302
37	37.4	3.6	47981	22	AAQ30757
38	37.2	3.6	355	21	AAQ74463
39	37.2	3.6	993	21	AAQ30765
40	37.2	3.6	993	21	AAQ30778
41	37.2	3.6	1616	20	AAQ33961
42	37.2	3.6	2260	19	AAQ41995
43	37.2	3.6	5335	21	AAQ66076
44	37.2	3.6	6225	20	AAQ55273
45	37.2	3.6	6225	21	AAQ20843

ALIGNMENTS

RESULT 1  
AAF82464  
ID AAF82464 standard; cDNA; 1031 BP.  
XX  
AC AAF82464;  
XX  
DT 29-JUN-2001 (first entry)  
XX  
DE Rat cDNA clone P00210D09.  
XX  
KW Rat; secreted factor; P00210D09; cardiatic; nephrotropic;  
KW antiinflammatory; gene therapy; cardiac disease; renal disease;  
KW inflammatory disease; ss.  
XX  
OS Rattus norvegicus.  
XX  
FH Key Location/Qualifiers  
FT CDS 96..923  
FT /tag= a  
FT /product= "Rat secreted factor"  
FT sig\_peptide 96..158  
FT /tag= b  
FT mat\_peptide 159..923  
FT /tag= c  
FT WO200123419-A2.  
XX  
PN 05-APR-2001.  
XX  
PD N.clavipes draglin  
XX  
PF Human transporter  
XX  
PF Human tumour suppressor  
XX  
PR Codon-optimised RA  
XX  
PR Triticum sp. cyste  
XX  
XX Maize caffeoyl-CoA  
XX  
XX Human TLE-2 gene.









CC (by viruses, bacteria, fungi, parasites, protozoa or helminths).  
 XX  
 SQ Sequence 9063 BP; 2492 A; 2546 C; 2654 G; 1371 T; 0 other;

Query Match 4.1%; Score 41.8; DB 22; Length 9063;  
 Best Local Similarity 55.0%; Pred. No. 1.7;  
 Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 QY 465 atcaaggctgcgtgaggttgccaaacccggtcttcacgtcttcctcgatgcc 524  
 DB 7943 acaagtcagtcgacgacgtttcaacgcccgcaggttcctctccagacgtg 8002  
 QY 525 cgtgccaaggactaccacaagaagtgcctcctgcagctcctgcagctgaagcagtcg 584  
 DB 8003 gatgacaagtatgaccgcatgaagactgctcctcctcagcagcagcagagccgcg 8062  
 QY 585 caggtggtcttcgtgactgggagtcg 613  
 DB 8063 gccctgaacgcgtgcagagtgagtg 8091

RESULT 8  
 AAV41730  
 ID AAV41730 standard; DNA; 1260 BP.  
 AC AAV41730;  
 XX  
 DT 20-NOV-1998 (first entry)  
 XX  
 DE Codon-optimised Ramy3D signal fused to DNA encoding mature AAT.  
 XX  
 KW Protein expression; monocotyledon plant cell;  
 KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;  
 KW ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;  
 KW antithrombotic; blood replacement; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.

Key Location/Qualifiers  
 misc\_feature 1..75  
 /tag= a  
 /note= "codon-optimised Ramy3D signal sequence"  
 misc\_feature 76..1260  
 /tag= b  
 /note= "encodes mature AAT"

W09836085-A1.  
 XX  
 PD 20-AUG-1998.  
 XX  
 PF 13-FEB-1998; 98WO-US03068.  
 XX  
 PR 13-FEB-1997; 97US-0038170.  
 PR 13-FEB-1997; 97US-0037991.  
 PR 13-FEB-1997; 97US-0038168.  
 PR 13-FEB-1997; 97US-0038169.  
 XX  
 PA (PHYT-) APPLIED PHYTOLOGICS INC.  
 XX  
 PI Rodriguez RL, Sutliff TD;  
 XX  
 DR WPI: 1998-467179/40.

XX  
 PT Expressing mature, glycosylated proteins in monocotyledonous plant  
 PT cells - from chimeric gene including signal peptide sequence,  
 PT specifically therapeutic agents and industrial enzymes  
 XX  
 PS Disclosure: Pages 34 iii-iv; 53pp; English.  
 XX  
 CC The present sequence encodes a fusion protein of codon-optimised Ramy3D  
 CC signal sequence/mature alpha1-antitrypsin (AAT). The protein is used

CC to exemplify the invention. The specification describes a method for  
 CC producing mature heterologous protein in monocotyledonous plant cells.  
 CC The method comprises transforming the cells with a chimeric gene  
 CC comprising a monocotyledon transcription regulator, inducible either  
 CC during seed maturation or by adding/removing a small molecule, DNA  
 CC encoding the heterologous protein, and DNA encoding a signal peptide,  
 CC with the signal peptide causing secretion of the protein from the cell.  
 CC Proteins expressed in this manner include mature glycosylated alpha  
 CC 1-antitrypsin (AAT) with a glycosylation pattern that significantly  
 CC increases its serum half-life, mature glycosylated antithrombin III  
 CC (ATIII), mature human serum albumin (HSA) having the native folding  
 CC pattern as shown by bilirubin-binding characteristics, or mature active  
 CC subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for  
 CC treating emphysema, ATIII as antithrombotic and HSA as blood replacement)  
 CC or as industrial enzymes (BPN' is used in detergents).  
 XX  
 SQ Sequence 1260 BP; 287 A; 428 C; 350 G; 195 T; 0 other;

Query Match 4.0%; Score 41.2; DB 19; Length 1260;  
 Best Local Similarity 47.6%; Pred. No. 1.4;  
 Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
 QY 464 catcaaggctgcgtgaggttgccaaacccggtcttcacgtcttcctcgatgcc 523  
 DB 759 catccagcactgcagaagctctccagctgggtgctcctcatgaagctctggggaacgc 818  
 QY 524 ccgtgccaaaggactaccacaagaagtgcgtcctgcagctcctgcagctgaagcagtc 583  
 DB 819 caccgcatcttcttcctgcgcgacgagggcagctccagcactggagaacgagctgac 878  
 QY 584 gcaggtggtcttcgtgactgggactgcgtgacgcacccaccctggctaccctggc 643  
 DB 879 gcacgacatcatcagaagttccttggaacgagagcagggcgctccgctagctccacct 938  
 QY 644 ttctgaggagatcgctccaccagcttctgccaagtgttcagctggaacagcagcaggt 703  
 DB 939 ccggaagctgagcatcaccgacgctgacgtgacgtgagagcgtgctgggccaagctg 998  
 QY 704 gtcggaggtgttaa 717  
 DB 999 caccgaagctctca 1012

RESULT 9  
 AAV82459/c  
 ID AAV82459 standard; DNA; 1747 BP.  
 XX  
 AC AAV82459;  
 XX  
 DT 16-MAR-1999 (first entry)  
 XX  
 DE Triticum sp. cysteine proteinase #4 encoding DNA.  
 XX  
 KW Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.  
 XX  
 OS Triticum sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 79..1497  
 FT /\*tag= a  
 XX  
 PN JP10327886-A.  
 XX  
 PD 15-DEC-1998.  
 XX  
 PF 27-MAR-1998; 98JP-0098140.  
 XX  
 PR 31-MAR-1997; 97JP-0114946.  
 XX  
 PA (SHOS ) SHOWA SANGYO CO.  
 XX  
 DR WPI: 1999-109255/10.











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 20:44:19 ; Search time 1988.65 seconds  
(without alignments)  
6843.926 Million cell updates/sec

Title: US-09-665-728-2\_COPY\_96\_920

Perfect score: 825

Sequence: 1 atgacgcctaggcgacgct.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_ov:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_om:\*

20: em\_or:\*

21: em\_ov:\*

22: em\_pat:\*

23: em\_ph:\*

24: em\_pl:\*

25: em\_ro:\*

26: em\_sts:\*

27: em\_sy:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htgo\_hum:\*

31: em\_htgo\_inv:\*

32: em\_htgo\_rod:\*

33: em\_htg\_hum:\*

34: em\_htg\_inv:\*

35: em\_htg\_rod:\*

36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Match	Length	DB	ID	Description
1	825	100.0	1031	6	AX108634	AX108634 Sequence
2	336.8	40.8	18207	9	AF156100	AF156100 Homo sapi
3	236.8	28.7	191750	2	AC024113	AC024113 Mus muscu
4	185.2	22.4	15597	3	AF074901	AF074901 Caenorhab
5	177	21.5	182211	9	AL360004	AL360004 Human DNA
6	171.2	20.8	191750	2	AC024113	AC024113 Mus muscu
7	171.2	20.8	230755	2	AC024618	AC024618 Mus muscu
8	108.4	13.1	41345	3	CEFI5G9	Z47068 Caenorhabdi
9	93.6	11.3	1610	9	HS245418	AJ245418 Homo sapi
10	91.4	11.1	4312	10	AF134319	AF134319 Mus muscu
11	89.4	10.8	157029	9	AL135796	AL135796 Human DNA
12	69.8	8.5	87461	9	HSG118G19	AL121996 Human DNA
13	69.8	8.5	180248	9	AC023275	AC023275 Homo sapi
14	65.4	7.9	717	11	HSJ34G10	AL159308 STS from
15	61.6	7.5	356	11	HS348TG1	Z51502 H.sapiens (
16	51.6	6.3	1873	3	NEUP37520	U37520 Nephila cla
17	50.4	6.1	2336	3	NCUPDSF	M37137 N.clavipes
18	50.4	6.1	2338	6	AR088543	AR088543 Sequence
19	50.4	6.1	2338	6	I92789	I92789 Sequence 1
20	48.6	5.9	87810	2	AC022987	AC022987 Homo sapi
21	47	5.7	4359	10	D85391	D85391 Mus musculu
22	47	5.7	41173	1	SCI41	AL132648 Streptomy
23	46.2	5.6	3867	1	AY033407	AY033407 Myxococcu
24	46	5.6	1026	3	AF350272	AF350272 Gasteraca
25	44.8	5.4	43349	1	SCF41	AL117387 Streptomy
26	44.2	5.4	44109	1	SC5F1	AL450165 Streptomy
27	44	5.3	12606	1	AE005995	AE005995 Caulobact
28	43.8	5.3	87810	2	AC022987	AC022987 Homo sapi
29	43.4	5.3	2745	6	AX179756	AX179756 Sequence
30	43	5.2	2047	10	BC003909	BC003909 Mus muscu
31	43	5.2	5991	10	AF199422	AF199422 Mus muscu
32	43	5.2	6423	10	AF199421	AF199421 Mus muscu
33	43	5.2	198677	1	AE001863	AE001863 Deinococc
34	42.6	5.2	132470	9	HS203P18	Z97180 Human DNA s
35	42.2	5.1	100000	9	AP000503	AP000503 Homo sapi
36	42.2	5.1	125350	2	AC020768	AC020768 Homo sapi
37	42.2	5.1	163682	9	HSDJ71117	AL132713 Human DNA
38	42.2	5.1	180283	9	AF134726	AF134726 Homo sapi
39	41.8	5.1	1682	1	STMGLNR	L03213 Streptomyce
40	41.8	5.1	11717	1	AE004509	AE004509 Pseudomon
41	41.8	5.1	32354	1	SCD84	AL353816 Streptomy
42	41.6	5.0	3203	10	AF008561	AF008561 Mus muscu
43	41.6	5.0	12356	10	AF134318	AF134318 Mus muscu
44	41.6	5.0	135545	10	MMHC213L3	AF109905 Mus muscu
45	41.6	5.0	224020	2	AC087117	AC087117 Mus muscu

ALIGNMENTS

RESULT 1					
AX108634	AX108634	1031 bp	DNA	PAT	30-APR-2001
LOCUS	Sequence 2 from Patent WO0123419.				
DEFINITION	AX108634				
ACCESSION	AX108634.1	GI:13923866			
VERSION					
KEYWORDS					
SOURCE					

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 1031)

Stanton, L.W. and Kapoun, A.M.

Differentially expressed genes

Patent: WO 0123419-A 2 05-APR-2001;

SCIOS INC. (US)

Location/Qualifiers

source

1. .1031

/organism="Rattus norvegicus"

CDS	/db_xref="taxon:10116" 96..923 /note="unnamed protein product" /codon_start=1 /protein_id="CAC37763.1" /db_xref="GI:13923867" /translation="MTPRAQLLPILLIATYTVAAAVTSDEPTKTLSPATGDTLAFVFDVTSMMDDLMQVLDGASRLILRSLSRVIANALVPFHPDIPGVLTADPVVFORELRLYVGGGDCPMSVGAIAKAAVEANPGSFIYFSDARAKDYHKNEILLQLLKQSOVFLVTGCGDPTHGYLAPEEIASTSSGOVPLDKQOVSEVLKVVESAIAQASKVHLLSADHEEGEHTWRIPFDSLKEVITLSLGGPPIEIVRDLGMSQSGSPPLMD"									
BASE COUNT	215	a	317	c	308	g	191	t		
ORIGIN										
	Query Match	100.0%	Score 825;	DB 6;	Length 1031;					
	Best Local Similarity	100.0%	Pred. No. 1.9e-152;							
	Matches 825;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	1	atgacgcctagggcgacgtcctcgcgtgtctcctcgtgacacctaacacagtagtgcgcg	60							
DB	96	ATGACGCTAGGGCGACGCTCTGCGCTGCTCTGCGGACCTACACAGTAGTGCGGCG	155							
QY	61	gcggtcacatctgatgagccacacgaagacgctgtccccgcacagagagacgcacctg	120							
DB	156	GCGGTACATCTGATGAGCCACGAAGACGCTGTCCCGCCGACAGGAGCGCCACCTG	215							
QY	121	gccttcgtcttcgatgtcacgcggctccatgtggagcagctatgacgtagtgcagcgc	180							
DB	216	GCCTTCGCTTCGATGTACCGGCTCCATGTGGGACGATCTGATGCAGGTGATCGAGCG	275							
QY	181	gcctacgcattctggagcgcagctctgagcagccgcgagcggtctatcgcaactatgcg	240							
DB	276	GCCTCACGATCTTGAGGCGCAGTCTGACGAGCGCGACGCGGCTCATGCCAACTATGCG	335							
QY	241	ctggttcctttccacgacccacagacatgtgcccaagtgaacctcagcgggacccagtggtg	300							
DB	336	CTGGTGCCCTTTCACGACCCACAGACATTTGGCCAGTGTAGCCCTCACGCGGACCCAGTG	395							
QY	301	tttcagagagagctgagacaactctatgttcaggggaggtggtgactgcccagaaatagct	360							
DB	396	TTTCAGAGAGAGCTGAGACAACTCTATGTTCCAGGAGGTGGTGACTGCCAGAAATGAGT	455							
QY	361	gtggggccatcaaggctgcggtgaggtgtgccaaccccggtccttcactcgtcttc	420							
DB	456	GTGGGGCCCATCAAGGCTGCGGTGGAGGTTGCCAACCCCGGCTCCTTCATCTACGCTTC	515							
QY	421	tcggaatgcctgtccaaggactaccacaagaagaatgagctcctgcagctcctcgcagctg	480							
DB	516	TCGGATGCCGTGCCAAGGACTACCAACAAGAAGAATGAGCTCCTGCAGCTCCTGCAGCTG	575							
QY	481	aagcagtcgcaggtgtgtcttcgtgctgactggggagctgcgtagcccgacccaccctggc	540							
DB	576	AAGCAGTCGCAGGTGGTCTTCGTCTGACTGGGCGACTCGGTGACCGACCCACCCCTGC	635							
QY	541	tacctggcttttgaggagatgcctccacgaagttctgtgccaagtgttccagctggacaag	600							
DB	636	TACCTGCTTTTGAGGAGATCGCTCCACCAAGTTCTGGCCAAAGTGTTCAGCTGGACAAG	695							
QY	601	cagcaggtgtcgaggtgttaaagtgggtgagtcgagtcgagcctcagggcctccaaagtcat	660							
DB	696	CAGCAGGTGTCGAGGTGTTAAAGTGGGTGGAGTCCGCCATCCAGGCTCCAAAGTTCAT	755							
QY	661	ctgctgtcagcagaccacgagagaggcggaacacacacacacacacacacacacacacacac	720							
DB	756	CTGCTGTACGACAGACACAGGAGGAGGCGCAACACACATCGGAATCCCTTTTGACCCG	815							
QY	721	agcttgaaggaagtaccatctcactgagcggcgagggcctgagatcgaaatccggagac	780							
DB	816	AGCTTGAAGGAAGTCAACATCTCACTGAGCGGGCGCAGGGCTGAGATCGAAGTCCGGAC	875							
QY	781	ccactgggttatgtcccaggtttcaactcctctcttgatgacaagac	825							

Db	876	CCACTGGGTATGTCCAGGGTTACCTCTCTTCTGTGATCAAGAC	920
RESULT	2		
AF156100			
LOCUS	AF156100	18207 bp	mRNA
DEFINITION	Homo sapiens hemicentin mRNA, complete cds.		PRI
ACCESSION	AF156100		
VERSION	AF156100.1	GI:14575678	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	1 (bases 1 to 18207)		
TITLE	Trent, J.		
JOURNAL	Human hemicentin gene		
REFERENCE	2 (bases 1 to 18207)		
AUTHORS	Trent, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Building 49 Room 4A22, Bethesda, MD 20892, USA		
FEATURES	Location/Qualifiers		
source	1..18207		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="1"		
	/map="1q24-q25"		
	230..17140		
	/codon_start=1		
	/product="hemicentin"		
	/protein_id="AAK68690.1"		
	/db_xref="GI:14575679"		
	/translation="MISWEVHTVFLFALLYSLAQDASQSEIRAEFEFPEGASTLAF VFDTGSMDDLVQVTEGASKILETSKREKRLFNFLFALVPHFIDPEGPVYITDPKK FOYELRELYVQGGDCPEMSIGAIKIALEISLPGSFYVFTDARSKDYLRTHEVLQI QKQSOVFLVLTGDCDDTHIGVYVEIASTSSGOVFLHDKQNVNKLWBEAVQA SKVLLHSTHLEQAVNTRIPFDPKLEKTVLSLGGSPMLIEIRNPLGKLKKGFLHE LLAHNSAKVYKPEAGMWTYKTSSSGRHSVRITGLSIDFRAGESRKTLDPKKT VSRVQGIPIYVLTSTPARDILLELSISGSRHSVRLTGLSIDFRAGESRKTLDPKKT DFVPEAFELKVTGDKDDYLFQVSSVSFSITVPDAPKVTPEKTPGYLQPGQIP CSDVSLPLFTLSEVRNGVTLGVQYLKESASVSLDAKATVLSDEGFECYFVAVSAGTG RAGTFDVSPPKVIQVNNVTVPGERAVLTGLIISAVDNLTWORNDRDLAEPA RIRTLANLSLEKSVKFNAGEVHCWVSSGSGSAASVLTVOEPKVTVMKNSPT GSEVSTMCATCYPKPKIAFTVNDMFIYCSHRYRMTDGTFLIKNAAPKADAGYGL ASNSAGTDKQNSLRYIEAPKLMVSELLVALGDTITVMECKTSGTPPPQVKKFGDL ELRPSFTLIIDPLGLLKTQETQDLDAGDYTCVAINEAGRATGKITLDVSGSPVFTOE PADVSMEIGSNVTLPQYVQGPETTKWRLLDNNPIFSRPFVSSTISQLRTGALFIL LWADSKGTICEAENQFKIQSETTVTGLVAPLIGISPSVANVTGGQTLPLCTIL AGNPERRWIKNSAMLLONPYITVRSDSLHTRVQLQDGGVTCVSNVAGTNNKT TSVVHVLPITHQOOILSTIEGIPVTLPCASGNPKPSVIMSKGELISTSAKPSA GADGSLIYVSPGESEGEYVCTATNAGAKRVQLTIVYRVRFGDLGSLGQDKPVE ISVLAGEVTLPCVKSPLPPIITWAKETQLISPPRHTFLPSGSMKITETITSDSG MYLCVATNAGNVTAQVKNLVHPPKIQRPKHLKQVQGVDPICNAQGTPLPVITW SKGSTMLVDGEHVSNDPCTLSIDQATPSDAGTYTCVATNAGTDETTLLHWQEP TVBDELPYNTTFQERVANORIEPCPKTPTIKLHNGRELTRGREGISILLEG TLVLASVTPYDNGEYICVANEAGTTERKYNLKVHVPVVKDKQSVSNVSLNQLT NLCEVEGTPSPILMMYKONVQTESTITQVNNKILKLFATPBDAGHYSCKALNI AGTSQKFIENIDVLVPTIITGNFPEVSVLVNDALECOVKGTPFDIHFWDKGPL FLDGNVELLDGRQVHLKLNARNDRKGYCTVSNAGKQAKDKITLYTPPSIKGN VTTDISLVNLSLIKLECKTRGLPMPAITWKDQPTMSSSOALYIDKQVHLHPRAQV VTDIDISLVNLSLIKLECKTRGLPMPAITWKDQPTMSSSOALYIDKQVHLHPRAQV SDSATVCHVANVAGTAESFHVDPVPMIEGNLATPLNKQVVIASHLTLECNAGN PSPILTWLDKGVVPKANDNRIEAGGKLEIMSQBEIDRGQYLCVATSVNAGEIKIYE VDVLVPPAIEGGDETSTFIVMNNLLELCHVTGSPPTIIMWLKQDGLIDGDFKIL LNGRLVIAQAVNSNTGLYRCMAANTAGDHKEFEFVTHVPTIKSSGSEVSVVYK PVALQIANGIPNPSTWLKDDQVNTAGNKLQSSGRVLTQATLLLEDAGTYTCVA TNAAGETQHIOLHHEPPLESADKNETVLVSNPVOLECAKAGNPVPIWTYKDN CLLSGSTMTFLNRGOI IDIESAQISDAGTYKCVATNSAGATELEFYSLQVHVAPSISG SNMVAVVNNPVRLECEARGIPAPSLTWLKDGSVSPSNGLOVLSGGRLLALTSTQ ISDTGRTCVANNAKEKQIDRLRVYVPPNMGEBEQNVSLISQVALLCQSDAIPP PITLWLKDGHPHLLKPLGLSISENRSVLKIEDAQVQDTRGTTCBATNVAGTEKKNNYV		
CDS			





JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

oriented line-shaped junctions  
Development 128 (6), 883-894 (2001)  
11222143  
2 (bases 1 to 15597)  
Vogel, B.E. and Hedgecock, E.M.  
Direct Submission  
Submitted (25-JUN-1998) Biology, Johns Hopkins University, 3400 N.  
Charles St., Baltimore, MD 21218, USA  
Location/Qualifiers  
1. .15597  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="X"  
/clone="cosmids F15G9; T09B9"  
1. .15597  
/gene="him-4"  
/note="mutants have defects in cell adhesion and germ-line  
chromosome segregation"  
1. .72  
/gene="him-4"  
1. .15597  
/gene="him-4"  
/function="cell adhesion"  
/note="alternatively spliced form has been detected"  
/codon\_start=1  
/product="hemikentlin precursor"  
/protein\_id="AAC26792.1"  
/db\_xref="GI:3328186"  
/translation="MGSRSPWLYGVGLLLLLATTCSSVDDKNDPTGKSLAFVDIT  
SKMYDVOVRGAAKIFKTVMAQREKLINYIMVPHDPYLGELTEINTDSTYFMQL  
GSVYHGDDGDEPKETLIKALQSLSPSFYIVFTDARKDYHLEDEVLNTIOERQS  
SVFVETGCGNRTHPGFTYERKIAAOSGVFHEKSDVSVLEVXRVHAKOKKVHL  
MYEARERTGYSKRNIPVDKHLSELATLSLGDKDDSDNLIDVLPDPSGTVDKRLYSKE  
GCTIDLVKVKLLRLKDPSPCVMTVNSRLKHTIRVFGHGVDFKVGFSRPLDRIEL  
ARPRVLQDITYLLINMTEGLIPGTVEGTEIDLVHGHSLYKAVASHPRTNPNMIFAGP  
FVPPGLFVRVQGDENDEMRATDAISVYGGVPRAFSPIHQFVGRDNLNLSL  
TVESASATYIYWKGTGDEIGGLPFTYHNTDSVMTPELSLKDAGEYECRVISNNQY  
SVKTYERRESPEIFGVRNVSPLGEAAFLHCSTRSAGEVELMRTRYCATVFNQNT  
ERNPTNGTLKIHVTRADAGVVECEARNAGMSTRKWLIDLRMPYSKVTPQDVYFNM  
REGVNLCEAMDGPKEVHHYFKGHHLLNDYKVGQDSKFLYIRUATHDEGTYECR  
ANSQAGARDTDLMLATPPKVEIIONKMWGRDVRFSFECKTRGKPHKIRWFKNG  
KDLIPDDYIKINEGLHIMGAKEADAGAYSCVGENMAGKQVANSVGRVPTIIES  
PHTVRVIERQYTLQCLAVGIPPELTIEWQKNVLLATLNNRQVTLADGNLLITDAQI  
EDQSGTCLARTNYQGQSGSTLMTVLGSPVLYGVHPPEOLETGGDLTSCVVLVLT  
PKPSTVIWKDKPVEEGPTIKTEGGSLRLRGKPNKDEGKYTCIAPSPAGNLSYV  
VOLIKKPEPVYKPEGIVKPTISGNDXKHVAVNSHDVLDGEGFAICPVWSGTPPP  
ITWLDGRPTPNRSRDFVTADNTLIVRKADSKSVGYTCQATNSAGONEQKTTIRI  
MNTPLSPQSGSNFMVVDLLTIPCDVYGDPKPVITLDDKPFTEGVNEDGSLTIR  
NYPNHRGFTYCEKJDSOLIPHTIRBDGALVLQNVLENTGVFCOVSNLAGEDSLY  
LPPAEARLTWYCEKJDSOLIPHTIRBDGALVLQNVLENTGVFCOVSNLAGEDSLY  
TLTVHEKPKIISEVPGVDVKGFTETIEPCRATGVPEVIRTNKNGIDILKMDKKFSV  
DNLGLRLIYEADKNDIGNVCVYTNAGTSQMTHTVDVQEPPIILSTQNTNNTAVVGD  
RVELCYVEASPSSVFWFRGAGNIDTKGYVESDGLTVLQOSASVEDATYTCRAS  
NIPADKAEANLWYLTASPKIDPDVVTDSYKESHPFSLYPCYFVSNLPQISWLNKD  
PLIDACTSNKWTGDKRLHVFRAKTDVSGYKCVARNAAGEGSKSFEVIVPLNLDL  
SKYKXVFAKEGEEVTLGCPGSGFPVQIMVYDVTVVBPBGKYKCATLSDGLTLHF  
DSVYKQEGNYHCVAQKNLIDVPELSVAVPIVGEDDNLNLEFVKDILSCDLQTLF  
ESDDKTYFWMSINGESDRPDNVQIPSDGHRLYITDAKENNGKYMCRVYNSAGKAE  
EAVFVNNHETVYVNVVTKYHTGVKCYATNKVQAAVKTINVHVRTKPRESGLTSESL  
TVNLTRSITLECDDDDAICGVISVYNGRPFPAETDGVQTLAGGERLHVHVASDTHGH  
STACTVITNAGVATITLNFVQVPTPIVNEGEGTYIENSLVLPCEVTGKPNVYTH  
TKDGRPVGLKSVLSEQQEQFKYHAEATHKGSYICMAKNMDVGTAEISFDVDITFRP  
MTQIKGKNTVAIKGALPKFCPIDDKNFKGQIWLNRVQPIDLEADARITRLNSD  
RLTLTILYNTDEGOYSCRVNHADENSDFKATVLPPTIIMLDKDKKTAVEHSTV  
TUSCATGKCPEDTFWDKDGEALHETIENADIIPNGELNGNLKITRIKEDGACKTYE  
AUNASGVEQDVNVNVTIPIKEGQIPDSYEQQONVCEVSCPVYATKAPATITWLKA  
GKPLASDQVTKSANGKLYLFKLRDTSKSKYTCIATNAGEDKDFKYSVWAPGSD  
ENPIVRRITVNSGNSPTLHCPAAGSPSTIWLKDGNAIETPNDRYVFDAGROLQYSP



QY 159 tctgatcgaggtgatgcagcggccctcagcattcttgaggcgcagctctgagcagccgcag 218  
 Db 150 TTGTGTCAGTGAAGGAGGTGCTGCTAAAATCTTCAAACATGTGATGCCAAGCTGA 209  
 QY 219 ccggggtcgcgcgaactatgcgtgctcttccacagccagacattggccagtgac 278  
 Db 210 AAGCTAATCTACACTACATCATGCTTCTTTCACGATCCATCTTAGGAGAGATTAT 269  
 QY 279 cctcagcggcggaccagtggttttcagagagctgagacaactctatgttcaggagg 338  
 Db 270 CAATACCACCGACTCAACTACTTTCATGAGACAACTCAGTAAGTGTACGTGCACGGAG 329  
 QY 339 tggtagctccagcaaatgagtgctgggggccaatcgaagctgcgctggaggttgcacaacc 398  
 Db 330 AGGTGATTTCAGAGAAGACGTTGACCGGTATTCTCAAGGCACCTTCAAATTTCTTACC 389  
 QY 399 cggctcctcatctacgtctctcagatcccgctgacagagcactaccacaagaatga 458  
 Db 390 GTCCCTCATTTATGTGTTCTCAGTATGCCCGATCAAGAGATTACCACCTTAGAAGATGA 449  
 QY 459 gctcctgcagctcctgcagctgaagcagtcgcaggtggtcttcgtgactggggaactg 518  
 Db 450 AGTTCTTAATAACAATTCAAGAGAAACAAGCTCAGTTGTTTTCGTGATGACTGGTACTG 509  
 QY 519 cggtagcgcagcccccctgctactcgtgctgttttgagagatcgcctccacagttctgg 578  
 Db 510 TGGTAATGAACACATCCAGGATTCGATACATATGAAATAATCGCGCTGCATCTTTTG 569  
 QY 579 ccaagtgtccagctggacaacagcaggtgctgcaggtggtttaaagtgggtgagtcgcg 638  
 Db 570 ACAAGTTTCCATCTTGAAAGAGTGAATGTCAGTACCGTATTAGATATGTCAGACATGC 629  
 QY 639 catcaggcctccaaagtctatctgctgcagcagccacagagggaggggcgcaacacac 698  
 Db 630 TGTCAAACAGAGAAGATTCTATTGATGTACGAGGACGCTGAACGGGGAGGAACAGTGTG 689  
 QY 699 atggagaatcccttttgaccacagcttgaaaggaagtccaccatctcactgagcgg 752  
 Db 690 TAGAACATCCCGTAGACAGCATCTCTCAGAGCTCACCATCTCATTGTTCAGG 743

## RESULT 5

AL360004 AL360004 182211 bp DNA PRI 16-AUG-2001  
 LOCUS Human DNA sequence from clone RP11-88G17 on chromosome 9, complete  
 DEFINITION sequence.  
 ACCESSION AL360004  
 VERSION AL360004.22 GI:15212049  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS Kimberley A.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 18, 2001 this sequence version replaced gi:15020463.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Emi., EMBL; Swi.,  
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr9  
 RP11-88G17 is from the library RPCI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6  
 This sequence is the entire insert of clone RP11-88G17 The true  
 right end of clone RP11-202H3 is at 75398 in this sequence.

## FEATURES

source 1..182211  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-88G17"  
 /clone\_lib="RPCI-11.1"  
 /3597..44209  
 /note="Sequence from AC006241 sequenced by WIBR"  
 50279..50351  
 /note="Sequence from AC006241 sequenced by WIBR"  
 93314  
 /note="Tandem repeat. Forced join. Gap sized to be  
 approximately 150bp from restriction digest data and  
 spanning pUC clone."  
 BASE COUNT 40368 a 47059 c 48127 g 46657 t  
 ORIGIN

Query Match 21.5%; Score 177; DB 9; Length 182211;  
 Best Local Similarity 78.2%; Pred. No. 9.3e-26;  
 Matches 226; Conservative 0; Mismatches 60; Indels 3; Gaps 1;  
 QY 1 atgacgcctagggcgcagctcctgcgcgtgctctctggcgacctacacagtagtgcgggcg 60  
 Db 137309 ATGATGCCGGGGCCGCTCCTCGCGTGTGCTGACCGGGTCTCTCGCGCAGTGCACGTG 137368  
 QY 61 gcgggtcacatctgatgagccacagagacgctgtctcccgccacagggagagacccacctg 120  
 Db 137369 GCAGTGGCC--GGGGCGCCCGGGACGGTAATGCCCGCCACACACGGGGGACGCCACCTG 137425  
 QY 121 gccttcgtcttcgatgcacggctccatgtgggacgatctgatgcaggtgatgcagcgc 180  
 Db 137426 GCCTTCGTCTTCGACGTCAACGGCTCCATGTGGGACGAACGTATGCAGGTGATCATGGC 137485  
 QY 181 gctcacgcattcttgagcagctctgagcagccgcagcgggttcacgcacacatcg 240  
 Db 137486 GCCTCGCATTTCTGGAACGCGAGTCTGAGCCCGCCGACGAGGCCATCGCCAACTACGCG 137545  
 QY 241 ctgggtgcctttccacagaccagacattggccagctgacccctacagcggcg 289  
 Db 137546 CTGTTGCCCTTCCACAGCCAGGTAGCGCCCCCGCCGACCCCGCCGCGCG 137594

## RESULT 6

AC024113/c AC024113 191750 bp DNA HTG 07-JAN-2001  
 LOCUS Mus musculus chromosome 11 clone RP23-277F6, WORKING DRAFT  
 DEFINITION SEQUENCE, 29 unordered pieces.  
 ACCESSION AC024113  
 VERSION AC024113.12 GI:12025591  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 191750)  
 AUTHORS Metzker, M.L., Lewis, J.R., Hume, J., Edwards, C., Harris, C.,





AC024618  
AC024618.2 GI:10280861  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 230755)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-22C1  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

2 (bases 1 to 230755)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campolano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collimore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wymann,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission

TITLE  
JOURNAL

Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 23, 2000 this sequence version replaced gi:7139711.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5311

Center clone name: 22\_C\_1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 206138 bases at least Q40

Consensus quality: 218628 bases at least Q30

Consensus quality: 223560 bases at least Q20

Insert size: 226355; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2864: contig of 2864 bp in length

\* 2865 2964: gap of 100 bp

\* 2965 4029: contig of 1065 bp in length

\* 4030 4129: gap of 100 bp

\* 4130 5611: contig of 1482 bp in length

\* 5612 5711: gap of 100 bp

\* 5712 6711: contig of 1000 bp in length

\* 6712 6811: gap of 100 bp

\* 6812 8073: contig of 1262 bp in length

\* 8074 8173: gap of 100 bp

\* 8174 9200: contig of 1027 bp in length

\* 9201 9300: gap of 100 bp

\* 9301 10629: contig of 1329 bp in length

\* 10630 10729: gap of 100 bp

\* 10730 12013: contig of 1284 bp in length

\* 12014 12113: gap of 100 bp

\* 12114 14005: contig of 1892 bp in length

\* 14006 14105: gap of 100 bp

\* 14106 15760: contig of 1655 bp in length

\* 15761 15860: gap of 100 bp

\* 15861 18390: contig of 2530 bp in length

\* 18391 18490: gap of 100 bp

\* 18491 20892: contig of 2402 bp in length

\* 20893 20992: gap of 100 bp

\* 20993 23366: contig of 2374 bp in length

\* 23367 23466: gap of 100 bp

\* 23467 26126: contig of 2660 bp in length

\* 26127 26226: gap of 100 bp

\* 26227 29160: contig of 2934 bp in length

\* 29161 29260: gap of 100 bp

\* 29261 33742: contig of 4482 bp in length

\* 33743 33842: gap of 100 bp

\* 33843 37207: contig of 3365 bp in length

\* 37208 37307: gap of 100 bp

\* 37308 41498: contig of 4191 bp in length

\* 41499 41598: gap of 100 bp

\* 41599 45456: contig of 3858 bp in length

\* 45457 45556: gap of 100 bp

\* 45557 50797: contig of 5241 bp in length

\* 50798 50897: gap of 100 bp

\* 50898 54710: contig of 3813 bp in length

\* 54711 54810: gap of 100 bp

\* 54811 59379: contig of 4569 bp in length

\* 59380 59479: gap of 100 bp

\* 59480 63217: contig of 3738 bp in length

\* 63218 63317: gap of 100 bp

\* 63318 66324: contig of 3007 bp in length

\* 66325 66424: gap of 100 bp

\* 66425 70575: contig of 4151 bp in length

\* 70576 70675: gap of 100 bp

\* 70676 75597: contig of 4922 bp in length

\* 75598 75697: gap of 100 bp

\* 75698 80005: contig of 4308 bp in length

\* 80006 80105: gap of 100 bp

\* 80106 86908: contig of 6803 bp in length

\* 86909 87008: gap of 100 bp

\* 87009 93267: contig of 6259 bp in length

\* 93268 93367: gap of 100 bp

\* 93368 97688: contig of 4321 bp in length

\* 97689 97788: gap of 100 bp

\* 97789 102663: contig of 4875 bp in length

\* 102684 102763: gap of 100 bp

\* 102764 120018: contig of 17255 bp in length

\* 120019 120118: gap of 100 bp

\* 120119 125200: contig of 5082 bp in length

\* 125201 125300: gap of 100 bp

\* 125301 131332: contig of 6032 bp in length

\* 131333 131432: gap of 100 bp

\* 131433 137931: contig of 6499 bp in length

\* 137932 138031: gap of 100 bp

\* 138032 146338: contig of 8307 bp in length

\* 146339 146438: gap of 100 bp

\* 146439 154167: contig of 7729 bp in length

\* 154168 154267: gap of 100 bp

\* 154268 161318: contig of 7051 bp in length

\* 161319 161418: gap of 100 bp

\* 161419 169637: contig of 8219 bp in length

\* 169638 169737: gap of 100 bp

\* 169738 178139: contig of 8402 bp in length

\* 178140 178239: gap of 100 bp

\* 178240 178339: gap of 100 bp

\* 178340 178439: gap of 100 bp

\* 178440 178539: gap of 100 bp

\* 178540 178639: gap of 100 bp

\* 178640 178739: gap of 100 bp

\* 178740 178839: gap of 100 bp

\* 178840 178939: gap of 100 bp

\* 178940 179039: gap of 100 bp

\* 179040 179139: gap of 100 bp

\* 179140 179239: gap of 100 bp

\* 179240 179339: gap of 100 bp

\* 179340 179439: gap of 100 bp

\* 179440 179539: gap of 100 bp

\* 179540 179639: gap of 100 bp

\* 179640 179739: gap of 100 bp

\* 179740 179839: gap of 100 bp

\* 179840 179939: gap of 100 bp

\* 179940 180039: gap of 100 bp

\* 180040 180139: gap of 100 bp

\* 180140 180239: gap of 100 bp

\* 180240 180339: gap of 100 bp

\* 180340 180439: gap of 100 bp

\* 180440 180539: gap of 100 bp

\* 180540 180639: gap of 100 bp

\* 180640 180739: gap of 100 bp

\* 180740 180839: gap of 100 bp

\* 180840 180939: gap of 100 bp

\* 180940 181039: gap of 100 bp

\* 181040 181139: gap of 100 bp

\* 181140 181239: gap of 100 bp

\* 181240 181339: gap of 100 bp

\* 181340 181439: gap of 100 bp

\* 181440 181539: gap of 100 bp

\* 181540 181639: gap of 100 bp

\* 181640 181739: gap of 100 bp

\* 181740 181839: gap of 100 bp

\* 181840 181939: gap of 100 bp

\* 181940 182039: gap of 100 bp

\* 182040 182139: gap of 100 bp

\* 182140 182239: gap of 100 bp

\* 182240 182339: gap of 100 bp

\* 182340 182439: gap of 100 bp

\* 182440 182539: gap of 100 bp

\* 182540 182639: gap of 100 bp

\* 182640 182739: gap of 100 bp

\* 182740 182839: gap of 100 bp

\* 182840 182939: gap of 100 bp

\* 182940 183039: gap of 100 bp

\* 183040 183139: gap of 100 bp

\* 183140 183239: gap of 100 bp

\* 183240 183339: gap of 100 bp

\* 183340 183439: gap of 100 bp

\* 183440 183539: gap of 100 bp

\* 183540 183639: gap of 100 bp

\* 183640 183739: gap of 100 bp

\* 183740 183839: gap of 100 bp

\* 183840 183939: gap of 100 bp

\* 183940 184039: gap of 100 bp

\* 184040 184139: gap of 100 bp

\* 184140 184239: gap of 100 bp

\* 184240 184339: gap of 100 bp

\* 184340 184439: gap of 100 bp

\* 184440 184539: gap of 100 bp

\* 184540 184639: gap of 100 bp

\* 184640 184739: gap of 100 bp

\* 184740 184839: gap of 100 bp

\* 184840 184939: gap of 100 bp

\* 184940 185039: gap of 100 bp

\* 185040 185139: gap of 100 bp

\* 185140 185239: gap of 100 bp

\* 185240 185339: gap of 100 bp

\* 185340 185439: gap of 100 bp

\* 185440 185539: gap of 100 bp

\* 185540 185639: gap of 100 bp

\* 185640 185739: gap of 100 bp

\* 185740 185839: gap of 100 bp

\* 185840 185939: gap of 100 bp

\* 185940 186039: gap of 100 bp

\* 186040 186139: gap of 100 bp

\* 186140 186239: gap of 100 bp

\* 186240 186339: gap of 100 bp

\* 186340 186439: gap of 100 bp

\* 186440 186539: gap of 100 bp

\* 186540 186639: gap of 100 bp

\* 186640 186739: gap of 100 bp

\* 186740 186839: gap of 100 bp

\* 186840 186939: gap of 100 bp

\* 186940 187039: gap of 100 bp

\* 187040 187139: gap of 100 bp

\* 187140 187239: gap of 100 bp

\* 187240 187339: gap of 100 bp

\* 187340 187439: gap of 100 bp







```
JOURNAL Immunogenetics 51 (4-5), 383-386 (2000)
MEDLINE 20260998
REFERENCE 3 (bases 1 to 4312)
AUTHORS Snoek,M., Van Kooij,A., van Vugt,H. and de Groot,K.E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Molecular Genetics, the Netherlands Cancer
Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands
FEATURES
    Location/Qualifiers
        1..4312
            /organism="Mus musculus"
            /strain="BALB/c"
            /db_xref="taxon:10090"
            /chromosome="17"
            /map="19.00 cM; between G7a and G7"
            /tissue_type="brain"
            1..4312
                /gene="G7c"
                /note="D17H6S56E-3; genomic structure of the G7c gene is
                presented in GenBank Accession Number AF134318; similar to
                the human sequence in GenBank Accession Number AJ245418"
                /allele="d"
                205..2880
                    /gene="G7c"
                    /codon_start=1
                    /product="G7c protein"
                    /protein_id="AAF61401.1"
                    /db_xref="GI:7381105"
                    /translation="MLPVEVPLSHLGLPILLQLQLPPTSAFFNWSLLAAGSVT
                    HDITTEALNIVTLVLEQPHGRPLRVEDYRGRTLLADDIFAAYPGPFSSRRFR
                    AALGSRANAQDFLPFAKSNPDLPDAERLVOCGRTRVLGALRRLTLVAARLEVTLA
                    RQRLGAALHALQDFYSHSNVWELGEROPHLLMRPROELWSLAQVDTCSDGSLSC
                    PGNMLDSTLLTSGYFGMHPAPPKCSHGHHFDQSSQPPRGKINDSTSPSPHHK
                    LHLQAEVALLASIAEAFSLRLGRDKAFSLRLDITPASSLFDLTGSGEEINAA
                    LIQARRIVEQSGPMPEVFIYLVFHDPGFVPFTTSDPDSFWQKLEIHALGGDE
                    PEMCLSALELALHTPFDIFVFTDASPKDALLNVRVSLTRRRCRVTLVETDPS
                    RTGRRRREALSPURFPEYEAIRASGEVIFTKDOYLQDVAAIVGSMAGLVTLPD
                    PVFTFGPCVFSVDSLLMOWTVRMHGDVSSFWIKSPAGVSGQPEEGIGPLCHTRRF
                    QFWITMDPTRTGTWELQVAAAGTPRVRVQAQSLDFLPHFGISVEDGPHGLYPLT
                    QPVAGLQLLVLTGTSRKLVGGOPOFHVLRVLRVPEGTQLGRVLSLEVPVPPVR
                    LLAASLPITLVSFPFSLGSGGSELRRTAPQCSVAPVLLLELSGPDFTLPG
                    SKAPLSHVSFSGPDILRTSNVPSFLTSLNLRARLGLNESAMGRLEWLEVPDAA
                    PDSVMTVTAAAGCAQVPPHTAFRLLLVLAQSSKQDLDPANSAAPVLPVPSALL
                    PSTLVTCAGGCGMAGRAWGTGVGLFLGCTSW"
                    1295..1524
                        /misc_feature
                        /gene="G7c"
                /note="exons around recombinational hot spot; corresponds
                to sequence presented in GenBank Accession Number
                AF008561"
                4254..4259
                    /polyA_signal
                    /gene="G7c"
BASE COUNT 830 a 1310 c 1166 g 1006 t
ORIGIN

Query Match 11.1%; Score 91.4; DB 10; Length 4312;
Best Local Similarity 52.4%; Pred. No. 1.7e-08;
Matches 225; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

QY 88 acgtgtgtccccgcagcagggagacccaccccttgcttctgttgatgtcacccggtccc 147
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1123 AGGCTGCTGGACATCATCCACCCAGCTTCCAGCCTGTGCTGCTGACACACAGGCAGT 1182
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 148 atgtgtggacgtctgatgcagtgatgcagcgcgcctcacgcattcttgagcgcagctgtg 207
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1183 ATGGCGAGGAATCAACGACGCCAAGATCCAGGCTGCCCGCATTTGGGACGACGCTCAA 1242
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 208 agcagccgcagccgggttcacgcaactatgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 267
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1243 GGCAGCCCCAT---GGAGCCTGTTTTCATATCTAGTGCCCTTCCAGACCCAGGGGTTT 1299
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 268 gcccagtgaccctcagcggcagccagctgggtgtgtgttcagagagagcgtgagacaactcat 327
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1300 GCCTCCGCTTTTACAAACAGCAGCCCGGACAGCTTTTGGCAGAAACTCAACGAGATCCAT 1359
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
QY 328 gttcaggaggtgtgtgactgccccagaaatgagtggtggggccatcaaggctgcggtggag 387
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1360 GCCTTGGGGGTGGAGATGACGACGAGATGTGCTGTCTGCTGAGCTAGAGCCCTGTTC 1419
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 388 gttgcacaccccggtctcttcacgtcttcacgtcttcacgtcttcacgtcttcacgtcttc 447
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1420 CACACCCCTCCCTCTCTGACATCTTTGTCTTCTACATGATGCTTACCAAGGATGCTCTT 1479
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 448 aagaagaatgagctctctcagctcctcgcagctcctcgcagctcctcgcagctcctcgcagctc 507
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1480 CTTACCAACGGGTGGAATCCCTGACTCGGAGAGCGCTGCAGGCTGACATTTCTAGTA 1539
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 508 actggggac 516
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1540 ACTGAAGAC 1548

RESULT 11
AL135796
LOCUS Human DNA sequence from clone GSI-164L12 on chromosome 1 Contains
DEFINITION an STS and GSSs, complete sequence.
ACCESSION AL135796
VERSION AL135796.6 GI:7706886
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157029)
Direct Submission
TITLE Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL requests: clonerequest@sanger.ac.uk
COMMENT On May 4, 2000 this sequence version replaced gi:7634143.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
GSI-164L12 is from the library Genome_Systems_ReleaseI VECTOR:
pBelobAC11
This sequence is the entire insert of clone GSI-164L12 The true
left end of clone GSI-15303 is at 103097 in this sequence.
Location/Qualifiers
    1..157029
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="GSI-164L12"
        /clone_lib="Genome_Systems_ReleaseI"
        1..798
            /note="L1P repeat: matches 1..791 of consensus"
            1064..1512
                /note="L1ME2 repeat: matches 5602..6150 of consensus"
```

```
misc_feature      complement(2521..2710)
/note="match: GSS: Em:AQ761942"
2792..3250
/note="match: GSS: Em:AQ246643"
3434..3489
/note="14 copies 4 mer tgt 87% conserved"
3435..3488
/note="27 copies 2 mer tg 88% conserved"
4119..4244
/note="MIR repeat: matches 23..152 of consensus"
6902..7098
/note="MER20 repeat: matches 8..218 of consensus"
7164..7191
/note="7 copies 4 mer ttat 96% conserved"
7500..7576
/note="L2 repeat: matches 2645..2739 of consensus"
7735..8041
/note="AluSq repeat: matches 1..308 of consensus"
9440..9901
/note="MLTLD repeat: matches 1..505 of consensus"
10595..10771
/note="MLT2 repeat: matches 277..553 of consensus"
10808..11111
/note="MLT2 repeat: matches 1..289 of consensus"
12089..12200
/note="L2 repeat: matches 2607..2749 of consensus"
12255..12361
/note="LIPB3 repeat: matches 6044..6150 of consensus"
12362..12675
/note="MER1B repeat: matches 1..337 of consensus"
12676..12916
/note="LIPB3 repeat: matches 5804..6044 of consensus"
13006..13612
/note="LIM1 repeat: matches 5180..5793 of consensus"
13614..13877
/note="AluJo repeat: matches 27..287 of consensus"
13879..13981
/note="LIM2 repeat: matches 2048..2147 of consensus"
13982..14292
/note="AluA5 repeat: matches 1..311 of consensus"
14293..15642
/note="LIM2 repeat: matches 640..2048 of consensus"
15608..15818
/note="LIPAL5-16 repeat: matches 146..337 of consensus"
15828..16373
/note="LIPBA repeat: matches -752..-208 of consensus"
16464..16704
/note="LIM2 repeat: matches -676..-428 of consensus"
16497..17042
/note="match: GSS: Em:AQ004976"
17721..19291
/note="L2 repeat: matches 1051..2708 of consensus"
19471..19528
/note="29 copies 2 mer aa 81% conserved"
complement(21071..21467)
21985..22024
/note="match: GSS: Em:AQ761948"
22384..22611
/note="MIR repeat: matches 3..262 of consensus"
22697..22863
/note="AluJb repeat: matches 137..302 of consensus"
23754..23861
/note="27 copies 4 mer acat 61% conserved"
24319..24381
/note="AluJb repeat: matches 84..148 of consensus"
24382..24670
/note="AluSq repeat: matches 22..309 of consensus"
24671..24822
/note="AluJb repeat: matches 148..289 of consensus"
26950..26985
/note="18 copies 2 mer tg 97% conserved"
26952..26987
```

---

```
/note="9 copies 4 mer tgt 97% conserved"
27687..27917
/note="AluJo repeat: matches 83..304 of consensus"
29345..29476
/note="L2 repeat: matches 2605..2750 of consensus"
30393..30828
/note="match: GSS: Em:AQ825928"
30990..31275
/note="AluSx repeat: matches 14..299 of consensus"
32321..32866
/note="match: GSS: Em:AQ418559"
33201..33344
/note="L2 repeat: matches 2562..2696 of consensus"
33363..33795
/note="L2 repeat: matches 2250..2690 of consensus"
35271..35548
/note="AluSq repeat: matches 18..297 of consensus"
complement(37223..37713)
/note="match: GSS: Em:AQ667379"
37319..37339
/note="match: STS: Em:G37800"
38212..38679
/note="LIMC1 repeat: matches 5847..6324 of consensus"
40470..40593
/note="MIR repeat: matches 8..140 of consensus"
41014..41123
/note="MIR repeat: matches 20..141 of consensus"
41395..41618
/note="MIR repeat: matches 10..356 of consensus"
43230..43494
/note="AluY repeat: matches 39..307 of consensus"
43683..44142
/note="LIMEC repeat: matches 1211..1677 of consensus"
44767..45295
/note="L1 repeat: matches 2530..3129 of consensus"
45533..45571
/note="AluJb repeat: matches 3..41 of consensus"
45604..46929
/note="LIPA7 repeat: matches 4758..6124 of consensus"
46930..47199
/note="AluJb repeat: matches 24..297 of consensus"
47358..47823
/note="L1 repeat: matches 3647..4127 of consensus"
47918..48310
/note="L1 repeat: matches 4257..4654 of consensus"
48311..48614
/note="AluSx repeat: matches 1..308 of consensus"
48615..48796
/note="L1 repeat: matches 4654..4834 of consensus"
49029..49516
/note="match: GSS: Em:AQ280129"
49087..49262
/note="LIP repeat: matches 5164..5357 of consensus"
49407..49510
/note="L2 repeat: matches 2579..2696 of consensus"
50481..51060
/note="MLTIF repeat: matches 1..536 of consensus"
complement(51651..52071)
/note="match: GSS: Em:AQ636105"
52290..52512
/note="match: GSS: Em:AQ797420"
52936..53296
/note="LIPAL5 repeat: matches 5785..6156 of consensus"
53303..53330
/note="7 copies 4 mer tgt 96% conserved"
53355..53457
/note="LIPAL5 repeat: matches 5689..5791 of consensus"
53680..53895
/note="108 copies 2 mer tt 57% conserved"
complement(54152..54622)
/note="match: GSS: Em:AQ568896"
54157..54274
/note="MER45B repeat: matches 199..318 of consensus"
```







repeat\_region 54316..54481  
/note="83 copies 2 mer tt 56% conserved"

repeat\_region 55309..55565  
/note="LIMD repeat: matches 1211..1474 of consensus"

repeat\_region 56353..56911  
/note="LIM4 repeat: matches 3491..4067 of consensus"

Query Match 8.5%; Score 69.8; DB 9; Length 87461;  
Best Local Similarity 63.3%; Pred. No. 0.00013;  
Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 101 ccacagagcgcaccctgaccttcgtccgatgcacggctcccatgaggcgacatc 160  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 38544 CCGAGGGGCGCTCACGTTCGCTTTTGTTGAATGCACGCTTCTATGATGATT 38603  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 161 tgaatcagtgcacggcgccctcacgccattctggagcgcagcttgagcagccagacc 220  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 38604 TAGTTCAGGTGATTGAGGGGCTCCAAAAATTTGGAGACGCTCTTTGAAAAGACCATAAA 38663

QY 221 gggatcgcaccaactatgcctggcgcttcccagaccagacattgg 269  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 38664 GACTCTTTCAACTTTTGCCTTGGCTTTCCATGATCAGGTAAAGG 38712

RESULT 13  
AC023275/c AC023275 180248 bp DNA PRI 09-MAY-2001  
LOCUS Homo sapiens BAC clone RP11-375L10 from 1, complete sequence.  
DEFINITION AC023275  
ACCESSION AC023275  
VERSION AC023275.3 GI:9838169  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998).  
MEDLINE 99063792  
REFERENCE 2 (bases 1 to 180248)  
. AUTHORS Swearngen,S., Maupin,R., Drone,K. and Gregory,S.  
TITLE The sequence of Homo sapiens BAC clone RP11-375L10  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 180248)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (10-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA  
REFERENCE 4 (bases 1 to 180248)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (17-AUG-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA  
REFERENCE 5 (bases 1 to 180248)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA  
REFERENCE 6 (bases 1 to 180248)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108,  
REFERENCE 7 (bases 1 to 180248)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108,  
COMMENT On Aug 17, 2000 this sequence version replaced gi:7630968.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0375L10  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-375L10; actual end is at base position 180248 of RP11-375L10.

#### FEATURES

source

1. 180248  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1"  
 /clone="RP11-375L10"  
 /clone\_lib="RPCI-11"  
 667..820

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region 12577..12788 /rpt\_family="Alu"  
 repeat\_region 12790..12957 /rpt\_family="Alu"  
 repeat\_region 13055..13209 /rpt\_family="L2"  
 repeat\_region 16593..17278 /rpt\_family="AcHobo"  
 repeat\_region 17344..17461 /rpt\_family="AcHobo"  
 repeat\_region 18161..18263 /rpt\_family="L1"  
 repeat\_region 18322..18687 /rpt\_family="L1"  
 repeat\_region 20558..21137 /rpt\_family="MaLR"  
 repeat\_region 22114..22321 /rpt\_family="L2"  
 repeat\_region 22353..22811 /rpt\_family="L1"  
 repeat\_region 22832..23003 /rpt\_family="L1"  
 repeat\_region 23004..23307 /rpt\_family="Alu"  
 repeat\_region 23308..23700 /rpt\_family="L1"  
 repeat\_region 23740..24219 /rpt\_family="L1"  
 repeat\_region 24419..24688 /rpt\_family="Alu"  
 repeat\_region 24689..26014 /rpt\_family="L1"  
 repeat\_region 26047..26106 /rpt\_family="Alu"  
 repeat\_region 26433..26642 /rpt\_family="L1"  
 repeat\_region 26654..28881 /rpt\_family="L1"  
 repeat\_region 27044..27199 /rpt\_family="L1"  
 repeat\_region 27370..27726 /rpt\_family="L1"  
 repeat\_region 28124..28388 /rpt\_family="Alu"  
 repeat\_region 29167..29275 /rpt\_family="MIR"  
 repeat\_region 30000..30223 /rpt\_family="MIR"  
 repeat\_region 30495..30677 /rpt\_family="MIR"  
 repeat\_region 31025..31217 /rpt\_family="MIR"  
 repeat\_region 32939..33406 /rpt\_family="L1"  
 misc\_feature 33434..33836 /note="similar to EST W86641 (NID:g1400389) zh63e05.s1"  
 misc\_feature 34177..34516 /note="similar to EST W86714 (NID:g1400443) zh63e05.r1"  
 repeat\_region 35831..35967 /rpt\_family="CRI"  
 repeat\_region 36070..36347 /rpt\_family="Alu"  
 repeat\_region 36643..36920 /rpt\_family="CRI"  
 repeat\_region 36969..37038 /rpt\_family="CRI"  
 repeat\_region 37823..38255 /rpt\_family="L2"  
 repeat\_region 38274..38387 /rpt\_family="L2"  
 repeat\_region 39802..39869 /rpt\_family="L2"  
 repeat\_region 40343..40628 /rpt\_family="L2"

```

FEATURES
  source
    1. 717
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /sex="male"
        /clone="SC9-12pJ34g10"
        /tissue_type="Lymphoblastoid cell line"
        /clone_lib="SC9-12pJ"
        186 a 198 c 211 g 120 t 2 others
        ORIGIN
          Query Match 7.9%; Score 65.4; DB 11; Length 717;
          Best Local Similarity 72.4%; Pred. No. 0.0036;
          Matches 84; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
          Qy 235 tatgcgtggctgttcctcacgaccagacattggccagtgacctcacgcgcgaccca 294
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
          Db 557 TCTGGCGTCCCTCTCTCTCTCTCCACAGATATTGGCNCAGTGACCTCACGGGGACCCC 498
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
          Qy 295 gtggtgttcagagagcgtgagacaaactctatgttcaggagggtggtgactgcc 350
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
          Db 497 ACAGTGTGTTTCAGAGGAGCCTCAGAGAACTTACGTGTCAGGTGGGAGCCCTTGACC 442
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 7, 2002, 20:47:59  
Job time: 7212 sec

[illegible]

Search completed: March 7, 2002, 20:47:59  
Job time: 7212 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 21:17:38 ; Search time 148.15 Seconds  
(without alignments)  
4774.170 Million cell updates/sec

Title: us-09-665-728-2\_copy\_96\_920  
Perfect score: 825  
Sequence: 1 atgacgcctaggcgagcgt.....ctcctcttctgatgcaagac 825

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_ll01.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	825	100.0	1031	22	Rat cdna clone P00
2	51.6	6.3	2004	18	Nephila clavipes s
3	50.4	6.1	2338	19	Nephila clavipes s
4	50.4	6.1	2338	21	N. clavipes spider
5	47.2	5.7	2338	12	N. clavipes draglin
6	43.4	5.3	2745	22	Human transporter
7	41.8	5.1	9063	22	Human tumour suppr
8	41.2	5.0	1260	19	Codon-optimised RA
9	40.8	4.9	1160	20	Maize caffeoyl-CoA
10	40.4	4.9	2271	15	Human TLE-2 gene.
11	40.4	4.9	2271	15	Human TLE-2 gene.

12	40.2	4.9	2214	19	AAV22682	New DNA sequence i
13	40.2	4.9	3331	19	AAV22683	New DNA sequence i
14	40.2	4.9	3331	20	AAZ32021	Human METH1 relate
15	40.2	4.9	3331	22	AAC90078	D86074 cDNA clone.
16	39.6	4.8	2502	16	AAO80752	DNA-polymerase con
17	39.6	4.8	2502	19	AAO65782	Thermus DNA polyme
18	39.6	4.8	2502	20	AAV63403	Consensus DNA sequ
19	39.4	4.8	4748	22	AAO06573	Bovine alpha(i) c
20	38.8	4.7	1929	19	AAV57472	Sorghum bicolor (L
21	38.8	4.7	2668	21	AAO00337	Wheat raffinose sy
22	38.6	4.7	1598	14	AAO40461	E.coli mdh promote
23	38.2	4.6	1995	12	AAO14184	N.clavipes draglin
24	38.2	4.6	1995	19	AAV23250	N.clavipes draglin
25	38.2	4.6	1995	21	AAZ38196	N. clavipes spider
26	38	4.6	1588	22	AAH26304	Spider silk protei
27	37.8	4.6	1747	20	AAV82459	Triticum sp. cyste
28	37.6	4.6	567	21	AAA29550	HIV codon altered
29	37.4	4.5	571	20	AAV89296	EST clone CH227.
30	37.4	4.5	3120	21	AAA12510	CDNA encoding a hu
31	37.4	4.5	5145	21	AAC98196	Human colon cancer
32	37.4	4.5	6085	18	AAV70153	S.longisporoflavus
33	37.4	4.5	31422	21	AAA92302	S. avermitilis ave
34	37.2	4.5	355	21	AAC74463	Human ORFX ORF18 p
35	37.2	4.5	5335	21	AAC66076	Human tyrosine kin
36	37	4.5	2000	18	AAV66543	50K-cellulase B ge
37	37	4.5	42000	21	AAA63349	Streptomyces globi
38	37	4.5	63164	21	AAA63348	Streptomyces globi
39	36.8	4.5	1886	20	AAZ32026	Human METH1 relate
40	36.8	4.5	1886	22	AAC90083	L23760 cDNA clone.
41	36.8	4.5	24379	18	AAV93095	Streptomyces freno
42	36.8	4.5	24379	19	AAV25925	Streptomyces roseo
43	36.8	4.5	50937	21	AAA09469	Streptococcus olea
44	36.6	4.4	426	7	AAAG0524	Sequence encoding
45	36.6	4.4	718	22	AAO08790	Human single chain

ALIGNMENTS

RESULT 1  
AAAF82464  
ID AAF82464 standard; cdna; 1031 BP.  
XX  
AC AAF82464;  
XX  
DT 29-JUN-2001 (first entry)  
XX  
DE Rat cdna clone P00210D09.  
XX  
KW Rat; secreted factor; P00210D09; cardiant; nephrotropic;  
KW antiinflammatory; gene therapy; cardiac disease; renal disease;  
KW inflammatory disease; ss.  
XX  
OS Rattus norvegicus.  
XX  
FH Key Location/Qualifiers  
FT CDS 96..923  
FT /\*tag= a  
FT /\*product= "Rat secreted factor"  
FT sig\_peptide 96..158  
FT /\*tag= b  
FT mat\_peptide 159..923  
FT /\*tag= c  
XX  
WO200123419-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 27-SEP-2000; 2000WO-US26582.  
XX  
PR 27-SEP-1999; 99US-0156277.  
XX  
PA (SCIO-) SCIOS INC.





XX Example 3; Fig 6A-D; 65pp; English.  
 XX The invention provides isolated cDNA molecules coding for spider silk  
 CC proteins. The spider silk proteins are characterized by repeating alpha  
 CC and beta regions and optional variable regions. The DNA sequences are  
 CC useful in the production of spider silk protein by recombinant DNA  
 CC techniques. The recombinant spider silk proteins may be used for the  
 CC production of fibers. The present sequence represents the cDNA encoding  
 CC the spider silk protein 1, derived from the major ampullate gland of  
 CC Nephila clavipes.  
 XX  
 XX Sequence 2338 BP; 564 A; 432 C; 918 G; 424 T; 0 other;

Query Match 6.1%; Score 50.4; DB 21; Length 2338;  
 Best Local Similarity 49.3%; Pred. No. 0.0088;  
 Matches 132; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
 Qy 419 tctcggatgcccgtgccaaagactaccacaagaagaatgagctcctcgagctcctgcagc 478  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 459 tcttggaaagcaaggtgcagagaggtggtattaggtggacaaggtgcagtgcaagcagc 518  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 479 tgaagcagtcagaggtgtctgtgactgggggactgcggtagcggtagcggtagcggtagcgg 538  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 519 agcagcagccggaggtgctggacaagggcgatagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 578  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 539 gctacctggtcttggagagatgcctccaccagttcttggccaaggtgttccagctgggaca 598  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 579 acaaggagctatgaggaacttgaagccaaggtgctggacagagagattagtgaggaca 638  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 599 agcagcaggtgtcggaggtgttaagtggtggagtcgcgcctccatccagcgcctccaaagttc 658  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 639 aggtgcaggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 698  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 659 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 699 acaaggtgctggacaagagcgtggagca 726

RESULT 5  
 AAQ14183  
 ID AAQ14183 standard; cDNA; 2338 BP.  
 AC AAQ14183;  
 DT 15-JAN-1992 (first entry)  
 DE N.clavipes dragline silk protein coding sequence.  
 XX protein superfibre; major ampullate silk; orb web spider; ss.  
 KW Nephila clavipes.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 1..2157  
 FT /\*tag= a  
 FT /product= silk protein 1  
 EP452925-A.  
 XX  
 XX 23-OCT-1991.  
 PD  
 XX 18-APR-1991; 91EP-0106217.  
 PF  
 XX 20-APR-1990; 90US-05111792.  
 PR  
 XX (UWY-) UNIV OF WYOMING.  
 PA  
 XX Lewis RV, Xu M, Hinman M;  
 PI WPI; 1991-312199/43.  
 XX P-PSDB; AAR14308.  
 DR

XX DNA encoding spider silk protein-1 and 2 and variants - isolated  
 PT from Nephila clavipes, for prodn. of spider silk protein and  
 PT fibres having desired characteristics  
 XX Claim 4; Page 23; 48pp; English.  
 PS  
 XX A N.clavipes major ampullate gland cDNA library was screened with  
 CC probes based on peptide fragments of the purified spider silk  
 CC protein (see AAQ14185). Positive plaques were identified and the  
 CC spider silk protein 1 coding sequence was determined. See also  
 CC AAQ14184.  
 XX  
 XX Sequence 2338 BP; 566 A; 433 C; 916 G; 423 T; 0 other;

Query Match 5.7%; Score 47.2; DB 12; Length 2338;  
 Best Local Similarity 48.5%; Pred. No. 0.051;  
 Matches 130; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
 Qy 419 tctcggatgcccgtgccaaagactaccacaagaagaatgagctcctcgagctcctgcagc 478  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 459 tcttggaaagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 518  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 479 tgaagcagtcagaggtgtctgtgactgggggactgcggtagcggtagcggtagcggtagcgg 538  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 519 agcagcagccggaggtgctggacaagggcgatagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 578  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 539 gctacctggtcttggagagatgcctccaccagttcttggccaaggtgttccagctgggaca 598  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 579 acaaggagctatgaggaacttgaagccaaggtgctggacagagagattagtgaggaca 638  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 599 agcagcaggtgtcggaggtgttaagtggtggagtcgcgcctccatccagcgcctccaaagttc 658  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 639 aggtgcaggtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 659 atctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 686  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 699 acaaggtgctggacaagagcgtggagca 726

RESULT 6  
 AAD09568  
 ID AAD09568 standard; cDNA; 2745 BP.  
 AC AAD09568;  
 DT 10-SEP-2001 (first entry)  
 DE Human transporter and ion channel-17 (TRICH-17) cDNA.  
 XX  
 XX Human; transporter and ion channel-17; TRICH-17; cystic fibrosis; mood;  
 KW gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;  
 KW hypertension; angina; neurological disorder; asthma; bipolar disorder;  
 KW dementia; depression; Alzheimer's disease; epilepsy; vaccine; arrhythmia;  
 KW Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;  
 KW Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;  
 KW demyelinating disease; mental disorder; Schizophrenia; polymyositis;  
 KW muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;  
 KW dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;  
 KW rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;  
 KW sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;  
 KW scleroderma; pulmonary artery stenosis; neutropenic; Addison's disease;  
 KW malabsorption syndrome; hypercholesterolaemia; cancer; ss.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 246..2066  
 FT /\*tag= a  
 FT /product= "Human TRICH-17 protein"  
 XX  
 XX WO200146258-A2.





Qy 490 caggtgtcttcgtgctgagtgaggactg 518  
Db 8063 gccctgaacgcgcgtgcagagtgagtg 8091  
RESULT 8  
AAV41730  
ID AAV41730 standard; DNA; 1260 BP.  
XX  
XX  
AC AAV41730;  
XX  
XX 20-NOV-1998 (first entry)  
XX  
DE Codon-optimised RAmY3D signal fused to DNA encoding mature AAT.  
XX  
XX Protein expression; monocotyledon plant cell;  
KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;  
KW AATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;  
KW antithrombotic; blood replacement; ss.  
XX  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH misc\_feature 1..75  
FT /\*tag= a  
FT /note= "codon-optimised RAmY3D signal sequence"  
FT 76..1260  
FT /\*tag= b  
FT /note= "encodes mature AAT"  
XX  
XX WO9836085-A1.  
XX  
XX 20-AUG-1998.  
XX  
XX 13-FEB-1998; 98WO-US03068.  
XX  
XX 13-FEB-1997; 97US-00381170.  
XX 13-FEB-1997; 97US-0037991.  
PR 13-FEB-1997; 97US-00381169.  
XX 13-FEB-1997; 97US-00381169.  
XX  
XX (PHYT-) APPLIED PHYTOLOGICS INC.  
XX  
XX Rodriguez RL, Sutliff TD;  
XX  
XX WPI; 1998-467179/40.  
XX  
XX Expressing mature, glycosylated proteins in monocotyledonous plant  
PT cells - from chimeric gene including signal peptide sequence,  
PT specifically therapeutic agents and industrial enzymes  
XX  
XX Disclosure; Pages 34 iii-iv; 53pp; English.  
XX  
XX The present sequence encodes a fusion protein of codon-optimised RAmY3D  
CC signal sequence/mature alpha 1-antitrypsin (AAT). The protein is used  
CC to exemplify the invention. The specification describes a method for  
CC producing mature heterologous protein in monocotyledonous plant cells.  
CC The method comprises transforming the cells with a chimeric gene  
CC comprising a monocotyledon transcription regulator, inducible either  
CC during seed maturation or by adding/removing a small molecule, DNA  
CC encoding the heterologous protein, and DNA encoding a signal peptide,  
CC with the signal peptide causing secretion of the protein from the cell.  
CC Proteins expressed in this manner include mature glycosylated alpha  
CC 1-antitrypsin (AAT) with a glycosylation pattern that significantly  
CC increases its serum half-life, mature glycosylated antithrombin III  
CC (AATIII), mature human serum albumin (HSA) having the native folding  
CC pattern as shown by bilirubin-binding characteristics, or mature active  
CC subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for  
CC treating emphysema, AATIII as antithrombotic and HSA as blood replacement)  
CC or as industrial enzymes (BPN' is used in detergents).  
XX

SO Sequence 1260 BP; 287 A; 428 C; 350 G; 195 T; 0 other;  
Query Match 5.0%; Score 41.2; DB 19; Length 1260;  
Best Local Similarity 47.6%; Pred. No. 1.2;  
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
Qy 369 catacaaggctgcggtgaggtgtcccaaccccgctctctcatctcttcggtatgc 428  
Db 759 catccagcactgcaagaagctctccagctgggtgctctcatgaagtacctggggaacgc 818  
Qy 429 ccgtgccaaagactaccacaagaagaatgagctcctcagctcagctcagctgaagcagtc 488  
Db 819 caccgcattcttctctcgcgacgagggaagctccacacctggagaaagagctgac 878  
Qy 489 gcaggtgtcttcgtgctgagtggtggtgactcggtgacgcacccacctggtacctggc 548  
Db 879 gcacgacatcatcaggaagtctctggagaaagcagagcgctccgctagctccacct 938  
Qy 549 ttctgaggagatgcctccaccagcttctggccaaagtgttccagctggacaagcaggt 608  
Db 939 ccggaagctgagcatcacgcgacgacgtacgacacctgaagagcgtgctggccagctgggcat 998  
Qy 609 gtcggaagtggttaa 622  
Db 999 cacgaagtgcttca 1012  
RESULT 9  
AAV25208  
ID AAV25208 standard; CDNA; 1160 BP.  
XX  
XX AAV25208;  
XX  
XX 19-JUL-1999 (first entry)  
XX  
XX Maize caffeoyl-CoA 3-O-methyltransferase cDNA.  
XX  
XX Maize; corn; caffeoyl-CoA 3-O-methyltransferase; lignin;  
XX transgenic plant; ss.  
XX  
XX Zea mays.  
XX  
XX Key Location/Qualifiers  
FH CDS 167..943  
FT /\*tag= a  
XX  
XX WO9910498-A2.  
XX  
XX 04-MAR-1999.  
XX  
XX 24-AUG-1998; 98WO-US17519.  
XX  
XX 12-MAY-1998; 98US-0076851.  
XX 27-AUG-1997; 97US-0057082.  
XX  
XX (PTON-) PIONEER HI-BRED INT INC.  
XX  
XX Bowen BA, Helentjaris TG, Wang X;  
XX  
XX WPI; 1999-204667/17.  
XX P-FSDB; AAY05669.  
XX  
XX Nucleic acids encoding plant lignin biosynthesis enzymes - used to  
PT transform plants to modulate lignin biosynthesis  
XX  
XX Claim 2; Page 138-139; 166pp; English.  
XX  
XX This DNA sequence encodes a caffeoyl-CoA 3-O-methyltransferase (see  
CC AAY05669) of maize. A polynucleotide having this sequence can be  
CC amplified from a cDNA library prepared from shoot culture of  
CC maize line CM45 using the primers given in AAX25241 and AAX25242. The  
CC invention provides methods and compositions relating to altering  
CC





DR P-PSDB: AAW56163.

XX New cDNA and e.g. vector, host cell and polypeptide - used to

PT produce polypeptide in high yields, which is used in cosmetics

XX Claim 2; Pages 7-9; 15pp; Japanese.

XX The present sequence represents a new DNA sequence isolated from Pinctada

CC fucata. The encoded polypeptide be used as an ingredient in cosmetics.

XX Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

XX

Query Match 4.9%; Score 40.2; DB 19; Length 3331;

Best Local Similarity 51.4%; Pred. No. 2.5;

Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 458 agctctgcagctcctgcagctgaagcagtcgaggtggtcttgcgtgactgggact 517

Db 766 agccgcagcgcgtgctgcctgcagggcgcaggtggaacttgaggactcgtggcg 825

Qy 518 gcggtgaccgcacccaccctggtacctggttttggaggagatcgctccaccagttctg 577

Db 826 acttgaggactcgtggtgagcacttgaggcctcgaggtctgtggtgctcgaggata 885

Qy 578 gccagtggtccagctggaacaagcagcaggtgctgaggtgttaagtgggtgagtcgcg 637

Db 886 tggaggatctgctgctgcctgctgtgctgcccgcgctgctgcccggaggaggact 945

Qy 638 c 638

Db 946 c 946

RESULT 14

ID AAZ32021 standard; DNA; 3331 BP.

AC AAZ32021;

DT 10-JAN-2000 (first entry)

XX Human METH1 related EST D86074.

DE Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;

XX cancer; diagnosis; hyperproliferative disorder; autoimmune disease;

XX angiogenesis inhibitor; abnormal wound healing; inflammation;

XX rheumatoid arthritis; psoriasis; endometrial bleeding disorder;

XX diabetic retinopathy; macula degeneration; haemangioma; detection;

XX arterial-venous malformation; immune deficiency; ss.

OS Homo sapiens.

XX

XX WO9937660-A1.

XX

XX 29-JUL-1999.

XX

XX 22-JAN-1999; 99WO-US01313.

XX

XX 23-JAN-1998; 98US-0072298.

PR 28-AUG-1998; 98US-0098539.

XX

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

XX

PI IrueLA-Arispe L, Hastings GA, Ruben SM;

XX

XX WPI: 1999-590684/50.

DR

XX New isolated metalloprotease thrombospondin polypeptides, useful for

PT treating hyperproliferative disorders, cancers or autoimmune disorders

PT

XX Disclosure; Page 322-324; 457pp; English.

XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human

CC metalloprotease thrombospondin (METH) proteins METH1 and METH2

CC respectively. METH1 and METH2 have been found to be potent inhibitors of

CC angiogenesis both in vitro and in vivo. They can be used for treating

CC cancer and other disorders related to angiogenesis including abnormal

CC wound healing, inflammation, rheumatoid arthritis, psoriasis,

CC endometrial bleeding disorders, diabetic retinopathy, some forms of

CC macula degeneration, haemangiomas, and arterial-venous malformations.

CC They may be useful in treating deficiencies or disorders of the immune

CC system, by activating or inhibiting the proliferation, differentiation,

CC or mobilisation (chemotaxis) of immune cells. The etiology of these

CC immune deficiencies or disorders may be genetic, somatic, such as

CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or

CC toxins), or infectious. They can also be used to treat inflammatory

CC conditions, both chronic and acute conditions. The products can also be

CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to

CC AAY49511 represent sequences given in the exemplification of the present

CC invention.

XX

SQ Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

XX

Query Match 4.9%; Score 40.2; DB 20; Length 3331;

Best Local Similarity 51.4%; Pred. No. 2.5;

Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 458 agctctgcagctcctgcagctgaagcagtcgaggtggtcttgcgtgactgggact 517

Db 766 agccgcagcgcgtgctgcctgcagggcgcaggtggaacttgaggactcgtggcg 825

Qy 518 gcggtgaccgcacccaccctggtacctggttttggaggagatcgctccaccagttctg 577

Db 826 acttgaggactcgtggtgagcacttgaggcctcgaggtctgtggtgctcgaggata 885

Qy 578 gccagtggtccagctggaacaagcagcaggtgctgaggtgttaagtgggtgagtcgcg 637

Db 886 tggaggatctgctgctgcctgctgtgctgcccgcgctgctgcccggaggaggact 945

Qy 638 c 638

Db 946 c 946

RESULT 15

AAC90078

ID AAC90078 standard; DNA; 3331 BP.

XX

AC AAC90078;

XX

XX 19-MAR-2001 (first entry)

XX

XX D86074 cDNA clone.

XX

XX METH; metalloprotease; thrombospondin; angiogenesis inhibition;

KW cancer therapy; benign tumour; ocular angiogenic disease;

KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;

KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;

KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;

KW coronary collateral; cerebral collateral; arteriovenous malformation;

KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;

KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;

KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;

KW Crohn's disease; atherosclerosis; birth control; ss.

XX

OS Unidentified.

XX

XX WO200071577-A1.

XX

XX 30-NOV-2000.

XX

```

PF 25-MAY-2000; 2000WO-US14462.
XX
PR 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX
DR WPI; 2001-025136/03.
XX
PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX
PS Claim 7; Pages 622-624; 768pp; English.
XX
CC The present invention relates to human METH1 and METH2, (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is an expressed sequence tag (EST) for METH. METH
CC can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer.
XX
SQ Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;

Query Match          4.9%; Score 40.2; DB 22; Length 3331;
Best Local Similarity 51.4%; Pred. No. 2.5;
Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 458 agctcctcagctcagctgaagcagctcagagtggtctgtctgactgggact 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 agccgcagccgctgctgcctgcagagggcagggctgtgaggaactcgtggcgg 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 518 gcggtgaccgcaccaccctggctacgtgcttttgaggagatcgccctccacagttctg 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 acttgaggagactcgtggcgacttgaggcctcgaggtctgtggtcctcgaggata 885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 578 gccaaagtgtccagtgacaaagcagcaggtgctcgagggtgttaaaagtgggtgagtcgg 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 tggaggatctgctgctgcctgctgctgctgcgcgcgctgctgcgcggaggtggagact 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 638 c 638
   |
Db 946 c 946

```

Search completed: March 7, 2002, 21:17:46  
Job time: 4559 sec